Identity and What Matters in Survival: Cause, Content or Continuity?

Andrew Barnes Rutherford

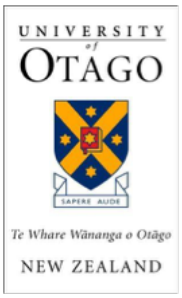
Submitted in fulfilment of the requirements for the degree of

Master of Arts, Philosophy

Philosophy Department

University of Otago

2021



> ZARABETH: Yes. He gave me weapons, a shelter, food. Everything I needed to live except companionship. He did not want it said that he had me killed. But to send me here alone, if that is not death, what is? A very inventive mind, that man. *Star Trek - "*All Our Yesterdays".[1]

Abstract

Derek Parfit maintains in *Reasons and Persons*[2] that personal identity consists in two things. Person A and person B are the same person only if:

1. A and B are psychologically very similar [R&P 216], that is, in their psychological states, including memory and desire, and also personality traits and tastes.

And:

2. A and B are connected by the "right kind of cause" [R&P 216]; that is, they have the same body over time.

Parfit also maintains, however, that personal identity does not matter, in the sense that what I care about when I have special concern for my future states does not depend on identity. Person B, really matters to Person A, in Parfit's view, only if:

1. A and B are psychologically very similar.

And:

2. A and B are connected by "any cause" [R&P 217]; that is, I may still care about a person who does not have the same body as me so long as we are very similar and connected in some other fashion. For example, a teleportation machine that destroys my body while reproducing the relevant psychological traits in another body is a cause that connects me to the other person in such a way that I will have special concern for that person.

---

[1] Shatner, William, Leonard Nimoy, DeForest Kelley, and Mariette Hartley. 1968. *Star trek, the television series. All Our Yesterdays.* Hollywood, CA: Paramount Pictures Corp. Accessed 13/11/20. http://www.chakoteya.net/StarTrek/78.htm

[2] Parfit, Derek. *Reasons and Persons.* Oxford: Oxford University Press, Incorporated, 1986. [Hereafter referred to as "R&P" and page number].
I invite the reader to use the following online resource to check my references to Parfit: http://www.stafforini.com/docs/Parfit%20-%20Reasons%20and%20persons.pdf Some of the page numbers are not the same as the original book but it allows easy searching for any particular piece of text.

I argue here that Parfit's second causal criterion is neither necessary nor sufficient for justifying special concern. I accept the first criterion as a necessary condition and introduce and defend a different second criterion that emphasises social connectedness. I maintain that A has justified special concern for B only if 1 obtains and:

2. Relevant social connections recognise A and B as the same person, and that person is in some way at least potentially accessible to them.

In defending this view, I present various accounts of personal identity and of what matters (Part 1 and 2), and I argue that Parfit's causal criterion for special concern fails (Chapter 3). Then, I present my account of social connectedness (Chapter 6) and defend its importance for justified special concern (Chapter 7). Finally, I give some examples that show how it works in various scenarios. (Part 3)

Acknowledgements

I want to acknowledge anyone whom my writing of this thesis has impacted, especially with it taking far longer than I intended. In particular, I would like to acknowledge the love and support of my family Simone, George and Sebastian.

Academically, I would like to acknowledge and thank my supervisors during this project: Michael LeBuffe, Charles Pigden and (briefly) Kourken Michaelian. All three have given valuable insights and advice that have improved the final product.

I want to say that any deficiencies in this thesis are entirely my fault and responsibility. If any academic shortcomings are noted it is almost certainly the case that supervisors have brought the issues to my attention, but I have chosen not to rectify them.

On the other hand, if there are any ways the thesis succeeds, it is entirely because of the teaching system at the Philosophy Department at the University of Otago.

Introduction

It is a curious fact that much of the important work done on personal identity has not been in volumes dedicated to the subject, but rather in works of philosophical genius that cover a broad range of subjects. From Locke, with An Essay Concerning Human Understanding,[3] to Parfit with Reasons and Persons. These works are central to understanding the general philosophical view of personal identity today, yet none focuses purely on it. This generalism suggests that advances in understanding, so far as personal identity is concerned, are accompanied by a broader philosophical understanding.

This work *does* focus on personal identity and the issues around it, but it also comes from a broader philosophical background. It was originally about the philosophical issues that arose from the movie *Groundhog Day* [4], which has become a classic by being the first significant and popular treatment of the metaphysical and ethical issues arising from being in a time loop. In the movie, Phil is a TV weatherman sent to a small town to cover the Groundhog festival along with his production crew. When he realises that he is trapped in a time loop, his behaviour becomes increasingly bizarre and antisocial. Despite the obvious change in his behaviour, his production crew treat him as being the same person. The movie's writers having characters ignore his change in personality is of interest because most of the major current theories of personal identity take psychological continuity to be the thing that connects our identity from one time to another. In this example, Phil does have internal psychological continuity but, (due to his behavioural change) from the perspective of the other protagonists in Groundhog Day he does not.[5]

Characters in Groundhog Day do not generally regard psychological continuity as important for personal identity at an intuitive level but are instead animalist. So long as his colleagues see Phil as being the same *animal* that they knew before, then he is also seen as being the same *person*.

Suppose the fiction of Groundhog Day is accurate in predicting general human responses when interacting with someone who does not seem to be psychologically continuous with their previous self. In that case, it suggests that we too are unconsciously animalists - when it comes to personal

---

Locke, John. *An Essay Concerning Human Understanding*. La Vergne: Neeland Media LLC, 2019. Internet resource.

[4] Ramis, Harold, Trevor Albert, C.O Erickson, Danny Rubin, Bill Murray, Andie MacDowell, Chris Elliott, Stephen Tobolowsky, Brian Doyle-Murray, George Fenton, Pe.mbroke J. Herring, and John Bailey. *Groundhog Day*. Culver City, Calif: Columbia TriStar Home Video, 2002

[5] I discovered that some of the most interesting issues in Groundhog Day were the ones of personal identity. Locke thought that personal identity was generated by memory.

Groundhog Day style thought experiments seem to come from a somewhat Lockean view of personal identity since the only difference between the aware and unaware participants is memory.

identity. Philosophers generally disparage animalism as a description of personal identity.[6] For instance, they generally take the brain to be a privileged organ when it comes to transplantation and would not advise its replacement. Animalism takes object identity (in terms of the animal) to be personal identity, whereas other views typically give the psychology of a person greater prominence. Perhaps psychological states are given greater prominence by people at an intellectual level because people are biased towards the intellect at an intellectual level.

Everyday people seem to be intuitively animalist, (as in seeing the identity of a person as being the presence of the same animal), whereby they take the markers of the animal identity of other people to be indicated by such things as physical appearance. Philosophers do not generally regard such things as appearance as being important in the intellectual understanding of personal identity. The difference between the popular and the philosophical understandings of identity suggests a contrast between immediate intuitions about personal identity and deeply examined ones. For instance, Phil in Groundhog Day is always treated by his production crew as he was before he entered the loop, regardless of the behavioural discontinuity which has obviously happened in him, due to his psychological changes from being in the loop.

Rather than being psychological continuity, intuitive animalism cashes out as Nozick's[7] "closest continuer" theory at an intellectual level, but there is more to be said about identity than just this. Our intuitions, generated from the fictional scenario seen in Groundhog Day, conflict with our supposed intuitions about identity in a number of different ways:

Firstly, in Groundhog Day, there is the incongruous fact about future survival, that a non-memory updating participant seems to be predicted to have less concern about being in a time loop, than the 'looping' person, who actually experiences ongoing psychological survival in normal terms. What this means is, that in fictional representations of time loop scenarios, characters who are not having their memories updated for each loop (like the central character is) might be predicted to be distraught, as it means a break in their psychological continuity and thus equivalent to our ordinary conception of death, whereas in fact they are depicted as being more at ease with the situation than the central character. This unconcern is the case even when they believe that the central character is in a time loop. This (again, if we are to generalise it to us) suggests that we have an intuitively 'majority rules' understanding of future survival. If there is the situation where the world is going to blink out of existence, to be replaced with a world as this one was a day ago, except for one person who will be inserted into that later world, then that is seen as a problem for that 'sole survivor' and not for the general population. This supposition supports a theory I later develop about social connectedness being important.

---

[6] Blatti, Stephan, "Animalism", *The Stanford Encyclopedia of Philosophy* (Fall 2020 Edition), Edward N. Zalta (ed.), https://plato.stanford.edu/archives/fall2020/entries/animalism/

[7] Nozick. Robert, *Philosophical explanations*, Cambridge, Mass. Harvard University Press, 1981. p. 29.

Secondly, Groundhog Day challenges our intuitions about causal history and how that should affect us, for instance by the presence, (or lack of) a build up of duties, responsibilities and debts from previous loops. I deal with examples of this at the end of the thesis.

I also came to the idea that background metaphysics (outside of our direct experience) should not be a determiner of how we should act instead of our apparent experiential relationships. I express this as the idea that "the information in the universe is the universe". A good example is Russell's five-minute world and my analysis of it.

As a sceptical hypothesis, Bertrand Russell imagines a situation where the world has only been in existence for five minutes. I interpret it as a metaphysical and ethical thought experiment, related somewhat to the Groundhog Day situation by also being a world in which history does not have its usual meaning, and it provokes us to examine what we should think about that. It became clear to me that Russell's five-minute world showed us that what we cared about in current and future states, was the current informational state and not prior causal states, and so it was of little importance whether these prior causal states ever existed.

I decided that a deep similarity (mostly of the psychological kind) was the sole determiner of identity, but at that stage I meant the narrow mental content within the person themselves, and any mental content identical to that, outside the person (the extended self). In order to help prove this view as the correct one, I became interested in the case of Swampman, and the literature on whether a randomly created being can produce meaning. I was interested in this because a randomly created being has no causal history.

"Swampman" is an example introduced by Donald Davidson[8], whereby a person enters a swamp and is hit, and vaporised by lightning. Then at the same moment, a different bolt of lightning hits nearby, and by some improbable occurrence, this second bolt creates an apparent person, physically and functionally identical to the original. The question is whether this closest continuer duplicate has thoughts and could generate meaningful statements? To me it was obvious that it did and could, and my question was whether it should inherit the same rights and responsibilities as the person who entered the swamp.

I believed that swampman could be a relevant example in the philosophy of personal identity as well. My idea was that since we would have to extend to a being that had no causal history, humanity, personhood, and other valuable traits, that would prove that causal history was not important to us. I was puzzled as to why Davidson had ever introduced the example in an attempt to prove the opposite thesis: That Swampman could create no meaning.

When I questioned visiting professor Sally Haslanger about her intuitions on the Swampman case if a Swampman replaced her husband Steven Yablo, I was very surprised that her intuitions

---

[8] Davidson, Donald. *Knowing One's Own Mind*, Proceedings and Addresses of the American Philosophical Association, Vol. 60, No. 3 (Jan., 1987), pp. 441-458, American Philosophical Association, http://www.jstor.org/stable/3131782

did not match mine, and became further fascinated by the philosophy of personal identity. I thereby changed my thesis to be focusing on personal identity, rather than repetitious situations.

Initially, my thesis focused on trying to prove that a causal link was not important in survival or personal identity, but rather similarity of mental content was. I came into contact with Parfit's ideas on personal identity, which have a causal link between two temporal stages and similarity between psychological content as some of the requirements, but it was far from clear what he meant by cause.[9]

Parfit has a two-stage process whereby he defines *personal identity* as psychological continuity and connectedness with the right kind of cause, but *what matters in survival* as psychological continuity and connectedness with *any* cause.

I agree with Parfit that what matters with respect to what we should care about is not identity but something else. Parfit takes that thing to be "Relation R" (psychological continuity and/or connectedness with any cause) [R&P 215]. I take what matters to be similarity between two spacetime points to as wide an extent as possible, including our current interpersonal relations, with a causal link not essential to survival.

Parfit says there is no "further fact" about personal identity, other than psychological continuity and/or connectedness with the right cause (and the right cause is *any* cause). More importantly, he says there is no further fact about whether a future person is worth caring about, other than psychological continuity and connectedness with any cause. My main disagreement with Parfit regards this 'further fact'. I agree that there is no further fact in the sense of a 'magical essence'. Yet, there is a further fact in what I initially called the 'extended person' as the identical traits we share with others, and later the 'network identity'[10], as the network of connections which we have with others. This social network is almost totally overlooked by philosophers studying personal identity, even though it is one of the main criteria by which ordinary people define identity. That is, the identity applied by others. I seek to prove that ordinary people are right in this instance.

In this way, I hope that I have met Professor Haslanger worry. A 'swamp Yablo' may not be as good as a regular Yablo, just because of social concerns about such things as randomly created beings. The social links may not exist between the swamp Yablo and other people in the same way due to them being viewed differently. This 'extrinsic' view is quite an outlier in the literature on personal identity but would be similar to the view that a deleted file is not as good as a regular file

---

[9] One of my main issues with Parfit at this point, is that many of his examples cause confusion and it is unclear exactly what he means. Late in writing the thesis I discovered in an appendix [R&P 477] a reply that Parfit had to Nozick regarding the theory that identity is via closest continuer rather than psychological continuity as Parfit says. I had always taken psychological continuity to effectively be the closest continuer in the psychological realm. Yet Parfit takes it to be something more like content similarity, which is what I had been trying to prove all along.

[10] Network might be considered to be social identity plus the identity to any entities that we might not consider to be part of social groups. e.g. computers.

on a computer. The fact is that when a computer 'deletes' a file, it does not actually erase it but rather tells the operating system to no longer regard it as existing within an index of files. So the content of a deleted file is the same as a regular file, yet it is how the network regards it that is of critical importance.

Later in writing the thesis, I came to two new understandings. Firstly, I believe there is no 'final fact' which determines identity in all conceptual cases. Parfit believes something like this when he says that:

> It is not true that our identity is always determinate. I can always ask, 'Am I about to die?' But it is not always true that, in every case, this question must have an answer, which must be either Yes or No. In some cases this would be an empty question. [R&P 216/217]

Since, under extraordinary circumstances, we may not be able to determine identity, and it is impossible to determine the exact boundaries of these circumstances, we can only say there are conventional facts about personal identity, but not final logical facts about personal identity. Thus, for instance, with regards to death, it is always contingent on future events whether death 'holds' as being the case.

The second realisation is to do with the psychological view of personal identity, which I call 'getting it' (at an intuitive psychological level as opposed to an intellectual level).

On 'getting it'

Parfit gives some examples that are supposed to cause the reader to reject intuitive personal identity. However, some of Parfit's examples are uncompelling. For example, when I originally read about the thought experiment of someone having each hemisphere of their brain transplanted into a different body, I was not convinced (in the way that I was supposed to be convinced) that there was a genuine problem to be solved. I could see a genuine problem to be solved in his teleportation case, but with split brains, Parfit did not convince me that people could really fission, by having their brain hemispheres separated, with two new people created. I presumed that it would result in two massively damaged persons.

It was not until I thought about death, and about the iceman found in an Italian glacier that I emotionally 'got it', in terms of understanding at a psychological level, what Parfit is saying about identity being indeterminate in certain circumstances. I asked: When does someone go from being a living being to being unretrievable in terms of that person's existence? My answer is that there is no specific point that this occurs. A person's existence does not depend on them having a certain level of psychological continuity, or even similarity. I have concern for my future self that graduates

from lesser to greater concern, depending on the level of similarity with my current self and current circumstances.

Let us imagine that we regard heart transplants as we typically regard brain transplants, and that we consider that a person 'dies' during the period in which the hearts get swapped over (as in 'I was clinically dead for five minutes during the surgery'). However, unlike our typical characterisation of that kind of clinical death, we apply the concepts of legal and social death to that understanding. That would mean that where someone says that they 'died for five minutes' (or similar), we would then start reading their legal will and treating them as a new person, much as people often suggest we do in the Swampman case.

Identity really 'gets real' when it comes to the issue of when we should apply the concepts of legal and social death. We are legally dead when others have the legal ability to act as if we are dead. We are socially dead when there is that kind of treatment socially. For example, people might think that Jeremy Bentham sought to avoid social death in some sense by having his stuffed body preserved and brought out for meetings at University College London. Yet, the question of *when* someone dies is secondary to the core issue. If I die and, then sometime afterwards, someone heaps up a lot of atoms and puts together a replica of me, then the intuitive conception of that is, it is all very well, but it does not help me. Yet intuitively, if my corpse gets resurrected, then it *does* help me. No one can tell where the line is between these things, such as if one has some of Jeremy Bentham's hair and builds the rest of his body around that. That is what makes personal identity as a concept fall apart, and that is what must be understood psychologically.

'Getting it' when it comes to personal identity is an all or nothing thing. It is almost a religious experience, and Parfit alludes to that, but because I am doing analytic philosophy, I will not emphasise this emotional part when analysing the issues.

Parfit says:

> I claim that a person is not like a Cartesian Ego, a being whose existence must be all-or-nothing. [R&P 275]

And:

> When I believed that my existence was such a [deep] further fact I seemed imprisoned in myself. My life seemed like a glass tunnel, through which I was moving faster every year, and at the end of which there was darkness. When I changed my view, the walls of my glass tunnel disappeared. I now live in the open air. There is still a difference between my life and the lives of other people. But the difference is less. Other people are closer. I am less concerned about the rest of my own life, and more concerned about the lives of others." [R&P 281]

If that doesn't seem particularly philosophical language, then the very last paragraph of Reasons and Persons says:

> The mental and material are really here, but here there is no human being to be found. For it is void and merely fashioned like a doll. Just suffering piled up like grass and sticks. [R&P 503]

Parfit gets this understanding of personal identity. What Parfit has not 'got', however, is full nihilism about personal identity. He still believes there is at least a graduated fact about personal identity, just that these facts are reducible to psychological continuity and the like. Parfit's philosophical conception of psychological continuity is a proxy for content similarity of a psychological kind. Parfit turns out to be gradualist about personal identity, which means that there can be no final answer about whether something has a particular identity - just its degree of similarity. Lack of facts is one of the things he does not go all-in on. The other is stating that content similarity is of importance in a blatant way. My view is that there can be a *further* fact - and that is network identity; there just cannot be a *final* fact. Like Parfit, I see the closest continuer view as the intuitive view of identity; however, it is not an absolute determiner of objective identity - since nothing could be, except where something is identical over time. Human beings are certainly not that.

Philosopher Robert Nozick originally put forth the closest continuer view of personal identity[11], whereby our identity over time is the being that is the closest continuer to us now. One basic problem with the closest continuer view is working out where the cut off is for when something is the close enough continuer, for there to be a supposed 'fact' about whether there is personal identity or not.

This means that there can never be a final fact about personal identity under all circumstances.

Conceptual considerations

It is worth pointing out that what follows is an intuitive, conceptual understanding of identity and what matters in survival. This means an understanding that makes sense in philosophical terms. I do not draw on the scientific facts of the matter very heavily, as this is a philosophic, not a scientific thesis. A scientific thesis might come to a completely different idea about identity, particularly regarding causality, as causality is seen as an important notion in science. Regardless of that, the main body of the thesis is on what matters in survival, and science should have a good deal less to say about that.

In terms of this topic, I feel I have made some very substantial progress. The reason for this is not so much the quality of my scholarship or level of intelligence. It is despite this. There is less research on this very important topic than might be expected, and much of what exists is fractured

---

[11] Nozick. Robert, *Philosophical explanations*, Cambridge, Mass. Harvard University Press, 1981.

and in silos. In particular, I think of the Swampman example being generally not examined by the philosophical community in terms of identity - but only meaning.

The major work in personal identity is by Parfit in *Reasons and Persons*, and the portion on identity is only part of a larger work with a broader agenda. Frankly, it is a work of genius, but there are some major deficiencies with it, which I will, of course, discuss.

More importantly, there has been no significant advance in the field since Parfit. Scott Campbell has made the most important advances, yet he is almost unknown.

The main way this thesis advances philosophical understanding, is to bring together disparate examples to more thoroughly refute *causal connection* as a determiner of identity and assign content similarity as the best determiner of what matters. Also important, is giving an extrinsic (external) role to networks in determining identity relations, and these have an element of subjectivity to them.[12]

Thesis summary

In science fiction and thought experiments, there may be multiple successor selves to our current self. Parfit, a major thinker on personal identity has said that our identity:

> Over time just involves (a) Relation R - psychological connectedness and or psychological continuity with the right kind of cause, provided (b) that this relation does not take a branching form… [R&P 216]

But

> Personal identity is not what matters. What fundamentally matters is Relation R with any cause. [R&P 217]

This thesis attempts to answer the basic question: What matters (in rational terms) regarding our future states? Is it maintaining the same identity, as per our intuitive understanding of identity, or is it something else? In order to answer this, we must see what creates identity relations, both intuitively and logically. Once we have done this, we can see if what matters is identity relations or something else.

I propose that what matters is 'content similarity'; That what matters about our future states is not whether they have a strict intuitive identity relation with us now, but rather whether there is a being

[12] That is, there is some element of thinking making it so.

that has maximum content similarity with us now, and this content similarity is as 'wide' as possible, so that it retains as many of the currently external connections as possible.

How are identity relations formed? What is the thing that means someone is the same person at another time or another place? Is it that there is a similarity between the people, or is it that there is a causal link between the people at one time and another?

There turns out to be no 'final fact' that can conclusively link persons at two separate times, as being the same person under all conceivable scenarios, but we typically use a heuristic; where identity relations are formed intuitively as:

1) That physical identity is a particular kind of spatiotemporal continuity formed by similarity at timeframes as short as we can possibly access, regardless of cause. And:

2) At the different times under consideration, there is a high degree of content similarity in the entity between one time and another. So we consider a person is the same being if they follow the same path through time and space. We will define the same person through time and space as the closest continuer person in content similarity to the already identified object/person in space/time. And:

The closest continuer continues to be *somewhat* similar to the initially identified entity so that it is not just the closest continuer by default, but rather by being increasingly similar at shorter intermediate timeframes.

Content similarity in relation to the identity of people/objects breaks down into two types:

One (the narrow) I define as the information inherent to the object/person, for instance (but not exhaustively) internal structure, as well as position and shape.

Two (the wide) network identity, would be defined as the extrinsic network of connections linking to the object from information held about it in other objects/people. However:

What we care about with regards to future states ('what matters') does not follow our intuitions about physical identity but is instead about content similarity being as close as possible to *present* content similarity (point to point). This similarity is important because it is as close as possible to what we care about now. This means that "our identity is not what matters", as Parfit put it, [245 R&P], but rather it is the future state of a person being as close to the present state of the person, to as wide an extent as possible. That being, that they retain as many of the links between them and other entities as possible and have their own internal similarity.

I have developed many arguments to support this view, from analysing (Davidson's) 'Swampman' scenarios to understanding how we would react in time-loop situations. I shall explain in detail how these scenarios are plausible, and what we can learn from them. All of this is regardless of cause.

People commonly believe causality to be key to identity and what matters. One basic argument I have against causality being important is that if a swamp-person appears, we would have no reason for treating it any differently to the original person due to its lack of causality alone. Also, in a time-loop scenario, it should not matter about the metaphysical background of why someone appears in the loop, but we should deal with them on the same basis as our previous interactions. For example: In the movie Groundhog Day, Phil is killed multiple times, and yet viewers still accept him as Phil (based on his content similarity) and would do regardless of the cause of his reappearance.

A lack of causality in the background metaphysics of events should not influence our view and have no implications for how we act. For instance, if the world came into existence recently, as in Russell's Five Minute World example[13], then that alone should not influence how we live in any way. Network identity is key to understanding why this is.

An explanation of terms

In this thesis, I talk about identity in intuitive terms, which means the practical terms we use to understand identity from day to day. This does not mean identity in absolute terms, because identity in absolute terms can only be between two instances of narrow content in an object that are identical even in space and time and are connected as being the closest identical arrangement. Technically, this means that identity in absolute terms is not possible for living beings whilst they normally remain alive. We can also talk about 'what matters' regarding survival, not in absolute terms but in terms of what is most 'reasonable' - Reasonable being; what we have most reasons to do, by it being most advantageous within our overall system of belief.

As for my term "network identity", it may be considered to be 'social identity *plus*'. That is social identity plus any other *relevant* wide content. A good example would be that a lot of information is stored about us in computer systems. If all the information about an individual was deleted we could say that this would not strictly affect their social identity but rather their broader network identity.

---

[13] Russell, Bertrand. *The Analysis of Mind*, The Floating Press, 2008. Ebook, p. 203. https://ebookcentral-proquest-com.ezproxy.otago.ac.nz/lib/otago/detail.action?docID=432353

On psychological continuity and/or connectedness

Throughout this thesis, I refer to Parfit's Relation R - psychological continuity and/or connectedness. Parfit defines psychological connectedness as "the holding of particular direct psychological connections" and psychological continuity as "the holding of overlapping chains of *strong* connectedness". [R&P, 206]
Parfit says this in his theory of personal identity before he goes into what matters about survival.

Parfit's concept of psychological connectedness is inadequate and somewhat misdescribes the supposed connection involved. When he says "connectedness", what that really amounts to is a similarity in states of beings, so that, if for instance, I have the same preference or belief as yesterday, I am considered to be psychologically connected to how I was yesterday. Parfit regards connectedness as inadequate to describe identity because someone, over their lifetime, is regarded as being the same person even though they may change a great deal and by that criterion have little 'connection' (meaning similarity) with how they were when they were young. Much later in Reasons and Persons [R&P, 301-302], Parfit considers whether an entity having continuity alone is sufficient for having what matters in survival. He regards it as not being enough - there must be 'connection' also. By *connection,* he means the same kind of psychological content, such as desires and memories.[14] To show this more clearly, he gives the example of immortal beings who change a lot over time and says:

> When such a person says for example 'I spent a period exploring the Himalayas' his hearers would not be entitled to assume that the speaker had any memories of the period, or that his character then and now are in any way *similar*… [My emphasis on similar]. [R&P 304]

Therefore Parfit's idea of connectedness and my idea of content similarity are much closer than I previously thought.

With regard to psychological continuity, it can't just be strong connectedness in the normal sense of 'connection', because you can have continuity without any *connection* at all. The word "connection" sounds like a link, but in the way that Parfit uses it is really used to say a shared similarity at the relevant times.
Parfit talks about continuity as "the holding of overlapping chains of *strong* connectedness" [R&P 206], but since I can't make much sense of this concept I prefer to think of continuity as: That which continues on in the same fashion as before, regardless of a specific link with what happened before.

---

[14] Also see: Bartels, D.M, and L.J Rips. "Psychological Connectedness and Intertemporal Choice." *Journal of Experimental Psychology: General*. 139.1 (2010): 49-69, p.50-51, http://home.uchicago.edu/bartels/ChoiceSymposium2013/02a-Bartels.pdf.

Why “content similarity” and “closest continuer”?

With regards to “content”, Parfit says that human beings have content and says that:

> Persons must be mentioned in describing the *content* [his emphasis] of countless thoughts, desires, and other experiences.” [R&P 226]

Content similarity is the concept of informational similarity within a specific bound. That being the boundary of the entity in question. I was asked, “Why not use the concept of properties? as in: Two objects are the same object if they have the same properties.”[15] The main reason is that *content* focuses on the substance of the object in question, whereas its *properties* can be almost anything related to the object, including extrinsic properties. Sven Bernecker says:

> What does it mean to say of two objects that they are similar? Intuitively, we say that objects are similar because they have most or many properties in common. But this intuition doesn’t take us very far, because all objects have infinite sets of properties in common.[16]

There is an issue of clarity when deciding whether to use ‘properties’ or ‘content’. If I refill my cup, it would be clearer to ask for the refill to consist of similar *content* to before, than to ask for a substance that has similar *properties* as before. It seems likely that a substance could have similar properties and yet be very different, whereas if it has similar content it will be more fundamentally similar. This is because similar content seems to entail more internally relevant properties than speaking of properties themselves.

An example of irrelevant extrinsic properties is that the only difference between objects might be that some people *think* they are different. In that case, it would be possible to say that there is a difference simply because some people falsely believe them to be different. The object/objects would have different properties but only the external property of being thought to be different.

The concept of properties breaks down into intrinsic and extrinsic properties. In my use of language, I call what might be called 'intrinsic properties' *narrow content* and extrinsic properties *wide content*. The terms ‘narrow’ and ‘wide’ content are usually used in relation to mental representational content, but I am using them to describe physical states of the world themselves. Narrow content refers to all that which is within the bounds of the person. The wide content includes the narrow, and also anything outside that to a particular scope. Where a person’s

---

[15] In conversation with Michael LeBuffe.

[16] Bernecker, Sven. *The Metaphysics of Memory*. Dordrecht: Springer, 2008, p.156.

relevant wide content involves information about them elsewhere in the world, I call it their network identity.

As for the name of the type of content similarity needed to link two spatio-temporally disparate objects together as being the same object/person over time, I originally used the phrase “Next closest” - as in our identity is maintained as the next closest thing to us through time and space. However, I found that Nozick was using “next closest” to denote the second closest object/person after the closest continuer. i.e.

> It may appear incoherent to care especially about the closest continuer qua closest continuer, but not care especially that there be a continuer that is closest. Focusing on the gap between the closest continuer and the next closest , it is as if I care a lot about this gap when it exists, but do not care, at all whether this is this gap.[17]

Because of Nozick’s usage, I thought that it would be too confusing to use “next closest” and would instead adopt the use of Nozick’s “closest continuer” labelling. However, this adoption is not without its problems because I am only talking about the closest continuer in terms of similarity, whereas Nozick is also talking about causal connection.

Nozick says:

> We care about continuation rather than merely that some qualitatively similar entity will exist… A later continuer, unlike a mere qualitative resembler, has its aspects connected, causally or subjunctively or however, to the characteristics of the earlier self; it would not be like it is (though that route) if the earlier self had not been that way. Part of the profile of the later self may stand to the characteristics of the earlier self in the very same relation a belief stands in to the fact when it constitutes knowledge. It is as if your later self knows you now, and so (except in special cases) views you as its closest predecessor. Connecting with that later self is a way of not sinking into oblivion. [Philosophical Explanations p.66]

This is a very poor argument, but that is what happens when people attempt to justify cause. Many philosophers believe that we need to have causality as part of identity. When trying to justify why we need causality, they say it is because we need causality. It is so bad, that it is literally the case that in the place where he is supposed to explain why causality is required, the famous philosopher puts a jumble of assertions and unconnected sentences. When is a future self you? Is Nozick saying it is when it is true that it is so, and when you are justified in thinking so? In that case it is self-referential. The defence of causality itself sinks into oblivion.

---

[17] Nozick, Robert. *Philosophical Explanations*. Cambridge, Mass: Harvard University Press, 1981, p.65.

On the identity of indiscernibles

Before talking about issues of when we have (or do not have) a certain identity, it is worth mentioning that there is an issue with assigning identity to any object or person that changes in any way over time, as, if 'two' things are identical in all respects, then they are one and the same. We can avoid this issue if we only refer to the narrow internal content of an entity remaining the same. However, with persons, even *that* narrow internal content is constantly changing, so what could form a stable means of assigning personal identity? This is why I say there is no final fact about identity.[18] I should also note that the philosophies of personal identity and object identity are different. In the metaphysics of the identity of objects in general, there is a great deal said which is not necessarily replicated in the literature on personal identity. It is not so much that the identity of objects is a superior overarching field but rather a field with different people working on it and different debates and elements. For this reason, it makes sense for me to focus more on personal identity than object identity.

Thesis structure

What this thesis sets out to achieve and where it achieves it.

The Introduction is the history of my thoughts on this topic.
The Thesis summary explains concepts in the thesis

Part one is on identity.
In chapter one, we establish what the candidates are for personal identity: Causal history or Content similarity?
Chapter two sees if we can remove cause from the equation, and I achieve this using the five minute world thought experiment. We also see if we can remove content from identity. We cannot remove content and still retain identity.

Part two is on what matters in future personal states.
We talk about identity as spatiotemporal continuity and then see how this works in terms of the Swampman argument, where cause is absent. We can see how the state of their content can give things meaning, as opposed to their prior relations (or lack of them). We can also see how by

[18] Noonan, Harold and Ben Curtis, *Identity*, The Stanford Encyclopedia of Philosophy (Summer 2018 Edition), Edward N. Zalta (ed.), URL https://plato.stanford.edu/archives/sum2018/entries/identity.

using wide content, we can overcome some of the problems that we have if we only use narrow content, thus ensuring that it is always content that is the determining factor, not cause.

The specific kind of wide content being advanced is called “network identity”. Wide content in general, and network identity specifically can provide a ‘further fact’ about our identity and what matters.

Part Three looks at some examples of theoretical situations such as time loops in order to test out theories of identity and what matters for survival under those conditions.

The findings of the thesis and their proofs

Because of the complexity and high level of understanding required to make sense of this thesis, I am much concerned that readers understand the ideas presented than I am in preserving the ideal academic format. Therefore, I provide this section as a basis to understand the claims I make and assess their plausibility.

Claim one: Causality is not essential in personal identity.

Academics typically see causality as extremely important in both science and philosophy to establish many things, including identity. However, the fact remains that when I remove causality from the picture, we are still inclined to see identity in the same way. That is, it is possible to describe identity without using causality, and so the concept seems to be redundant. I do not know the extent of the redundancy beyond the study of identity. Because causality is generally an assumed claim, there is often little to argue against. I do my best to challenge arguments for causality in part one when Shoemaker argues for the inclusion of causality as an essential prerequisite for identity. The claim that causal connection is not required for identity is an important one, but not the main thrust of this thesis.

Claim two: A causal connection between their existence at two different points is not essential for a person having what matters in survival.

Parfit draws a powerful distinction between the identity of a person and that person having what matters in survival. Parfit convincingly claims that what matters is not preserving an individual's identity but rather their psychological continuity and connectedness. Parfit is able to do this by comparing our preservation of identity with other normal physiological process which we find important and which can be substituted by other means of achieving those processes. For instance, he says about someone who uses an artificial eye instead of their natural eye:

> Would this person be seeing these objects? If we insist that seeing must involve the normal cause, we would answer No. But even if this person cannot see, what he has is just as good as seeing, both as a way of knowing what is within sight, and as a source of visual pleasure. If we accept the Psychological Criterion, we could make a similar claim. If psychological continuity does not have its normal cause, it may not provide personal identity. But we can claim that, even if this is so, what it provides is as good as personal identity. [R&P 209]

I accept Parfit's claim, and my further claim is that if there is no causal connection, then so long as there is psychological continuity and connectedness, there is what matters in survival. Similarly, if a blind person is hallucinating, if that hallucination is veridical with what they would otherwise be seeing, then that hallucination is as good as seeing, so long as that hallucination accurately portraying events remains the case. An example that I use is that if someone puts a coin into a drink machine and the machine malfunctions so that it does not recognise the coin, yet accepts it and simultaneously malfunctions so that it dispenses a drink; then the coin user is the owner of that drink regardless of the lack of a causal chain.

This would be wonderful and easy to present as my main thesis, however a philosopher called Scott Campbell got there first. He imagines the situation of teleportation, and a machine holding information digitally about a person during the process of transportation. If this information is substituted for random information, then it would not threaten what matters about survival so long as the information is identical. Campbell calls this view "Psychological Sequentialism"[19].

Regardless of finding that I am not totally original in promoting a 'sequentialist' viewpoint, causality's lack of importance in survival is nonetheless useful in helping to establish the main argument of my thesis. That is, that what matters in survival is more than just the individual's connections to themselves (in the Parfitian sense, of there not needing to be an actual connection) but also the individual's connection to the world.

Claim three: Extrinsic (external) connections to the world are an essential part of what matters in individual survival.

This claim is the most important one that I make because it is the most original. That said, I am not the only person to claim that extrinsic properties matter in survival. A philosopher called Heidi Savage also claims this; however, I believe her proofs are more dubious. [See Chapter 5]

In Parfit's explanation of what matters, connections between the person at different times cause the person to have concern for their later self. The 'connections' are in fact a similarity of mental content between the two times, as opposed to 'linked' connections as we usually understand

---

[19] Campbell, Scott. "Is Causation Necessary for What Matters in Survival?" *Philosophical Studies*. 126.3 (2005): 375-396. [Hereafter “Campbell”, date and page number]

connection. In the case of external connections it is a link in the sense of another person referring to that person and information being able to past between them - causally or not.

A more succinct way to understand where Parfit has gone wrong is to see that he has imported an extra concept in addition to what he has said is necessary for survival into his transporter example. The transporter replicates a consciousness, which is the thing that Parfit says is important, and it replicates a body. If it was just to transport a consciousness and leave the body, such as in the Christian concept of the afterlife, but also created the situation that in addition to the Christian concept, the consciousness could not interact with anything, then that would be effectively equivalent to death even though it would fully meet Parfit's criterion for life.

Part 1 - What does identity consist of?

Introduction

Parfit has a view of what personal identity consists of, and that is: “Psychological connectedness and/or continuity with the right kind of cause” [R&P216]. Yet, he mainly discusses the views of other writers and rarely clearly argues for his own view of what identity and personal identity are. It seems that the primary purpose of Parfit’s discussion of personal identity is to establish the ground from which to move to talking about the thing he is mainly interested in, which is what matters in survival. That is: Parfit is merely describing personal identity in order to say that personal identity doesn’t matter; what matters is psychological connectedness and/or continuity with any cause.

Parfit has two additions to his theory of identity:

One is identity not taking a *branching* form, so that there cannot be two of the same person.
The other is that it has the right kind of cause.
Of these, only “cause” makes its way into the account of what matters.
Parfit makes no specific argument for the inclusion of cause into this theory of personal identity. Instead, he says: “I follow Shoemaker” [R&P 516] in reference to an argument about the casual dependence of memories. [R&P 220].
Parfit argues that conceptually we can only remember our own memories; otherwise, they are not defined as memories. Parfit then introduces the idea of “quasi-memories” solving the definitional problem and says they are veridical where:

> (1) I seem to remember having an experience,
> (2) I did have this experience, and
> (3) My apparent memory is causally dependent, in the normal way, on this past experience. [R&P 207]

Because Parfit does not specifically set out a theory of why cause is important, and says he follows Shoemaker’s theory of why it is important, I look at Shoemaker’s theory of cause in his theory of identity. Shoemaker also says similarity needs to be connected by cause and that similarity alone is not enough to have a theory of identity. Because I am directly arguing that similarity over time *is* enough, that is another reason to engage with Shoemaker as opposed to Parfit at this point.

Chapter 1 - Candidates for identity description

When talking about identity, we can see that there is often a difference in criteria between object or entity identity, and personal identity. As a starting point for discussion, personal identity should follow the object/entity identity of the substance that the person consists of. This is until there is a reason to break them apart, which is what we see when people give psychological states a greater level of importance than their other bodily states.

For general theories of identity, there are two likely candidates:

The *Causal connection* theory: Current entities and objects can only be the same entities and objects at a later time, as a result of the effect of the previously existing entities and objects making them be the way that they are. Cause both precedes and explains subsequent effects.[20]

The *Content similarity* theory: Current entities and objects are identical with previous entities and objects by having deep similarity with earlier entities and objects.

It may also be claimed that both causal connection *and* similarity are required.

The argument against causality in identity

P1: When assigning a particular identity to entities, “causal connection” and “content similarity” are the two most promising candidates for providing the descriptive role.

P2: We can deprive entities and objects of causal connections without changing our description of their identity.

P3: We can not deprive entities and objects of similarity without changing our description of their identity.

Therefore: The similarity of entities and objects must be the thing that is of importance in describing their identity.

What is causal connection? A causal connection exists if there is a link between two states, such that the earlier state made the later state happen, or one entity is producing some currently efficacious force upon another entity, such as a person holding down a spring.

---

[20] There is one possible exception and that is backwards time travel, in which case a future cause is supposed to cause a past event.

What is content similarity? There are three kinds of content similarity that I talk about in this thesis:

1. Similarity in narrow content: The internal structure of a particular person or object, such that if those structures are duplicated, there is no way to tell between a copy and the original, apart from spacial position.

2. Similarity in wide content: Similarity in the world outside the object or entity in question.[21]

3. Similarity in relational or network content: A subset of wide content forms what I call network identity. Network content that exists in relation to a particular person or object, I call its network identity. i.e. its network links and the information held about it in other entities.
I examine network content later in the thesis.

Similarity *and* spatio-temporal continuity

My formulation of what content similarity is, in relation to identity over time, is that an object's identity is formed by it being the thing that is most similar to it previously, both in spatio-temporal position and in internal structure. That is, that during the next time slice, the narrow content of an object is as close as possible to the original object and the spatio-temporal position is as close as possible. These two criteria together form *the closest continuer theory*. I posit the closest continuer theory as our basic intuitive understanding of identity.

Similarity *and* spatio-temporal continuity *and* cause: Shoemaker

There are other candidates for what forms identity. For example: Similarity *and* spatio-temporal continuity *and* cause. That is: Something is the same thing if it remains similar, while it follows the same spatio-temporal path, and the later parts of the spatio-temporal path are linked to the earlier, by those being the ones that cause the later parts to be as they are.

---

[21] Identity is often seen as both causal connection and content similarity. For instance, under US copyright law there is a copyright liability where a person has access to a song and produces a similar song.
So it is both *potential* cause and actual similarity. This means under law that identity is determined by both cause and similarity. To be clear, under the law it is only whether or not a *copy* has been made in some way that is the question. i.e cause. However, when *determining* this, it is similarity that comes to the forefront. There can be two identical works by different authors and different rights pertaining to them so long as the chains of causality don't intersect.

In *Identity, Properties and Causality*, Sydney Shoemaker argues for a view of identity as spatio-temporal continuity linked by the right sort of causal dependence.[22] I accept that identity is spatio-temporal continuity, but reject the need for causality.

Shoemaker tries to defend the causal connection theory, whereas many other philosophers just presume it to be true. Also, Shoemaker's arguments for causality are at least coherent, whereas Nozick merely asserts that causality is required.[23]

In his argument for the necessity of causality, Shoemaker first of all considers what might be called a 'naive' version of content similarity, without spatiotemporal continuity. He says, "Similarity is not always good evidence of identity." [Shoemaker 324]. Shoemaker imagines a group of chairs painted in various colours. Every so often, these chairs are randomly painted other, different colours. By repainting the chairs, there is no link between similarity in the colours of the chairs and their identity as the same chair. In the case outlined, the disconnect is between the identity of the colours and the identity of the chairs. Their identity in terms of colour has changed, but not in terms of being the same physical chair. His point, is to move his argument from being about *apparent* similarity to being about *spatiotemporal* similarity, which is a necessary refinement of the similarity thesis.

Shoemaker has a second argument for the necessity of causality. He images a machine that causes a table to disappear and a second machine that causes a table to be produced in the same place. If the machine happens to have the same settings for the production of the new table as it did for the old one, then there would be a spatiotemporally continuous identical table, but it would be a different table. [Shoemaker 326] There are many things that I could say about this.[24] For example, if there was a machine that could instantly replace one table with another, we would need to have a different understanding of identity, but as it is we don't. Yet, I do not need to use such an argument with either the tables example or Shoemaker's next theological example with God and stone tablets.

If we put the argument in a logical form, we can see what is wrong with it:

1. There is a table.
2. The table disappears.
3. *Instantly,* after the table disappears, another, *different* table, takes its place.
4. We know that a different table has taken its place.

---

[22] Shoemaker, Sydney, *Identity, Properties and Causality*, Midwest Studies in Philosophy, 4, 1979 p. 321-42. [Hereafter "Shoemaker" and the page number]

[23] Nozick, Robert, *Philosophical explanations*, Cambridge, Mass. Harvard University Press, 1981, p.35, https://antilogicalism.com/wp-content/uploads/2017/07/philosophical-explanations.pdf

[24] I could use the trick that teleosemantics uses to dismiss Swampman (or at least David Papineau uses when he said "Merely Possible Cases are Irrelevant").

5. The different table is spatiotemporally continuous with the previous table.

Therefore: 'There must be more to identity than spatiotemporal continuity'.

Shoemaker defines the tables as different due to their having a different causal history, whereas when this claim is examined in detail, the example does not work as we shall see. For something to have spatiotemporal continuity, there must be no time slice so small, that when the original object disappears, the new object is not there to follow on from it. This is a fundamental problem for the table example because although it seems to work at first, it is not possible to imagine at an atomic level.

There are two problems with the table example:

First of all, there cannot be a time period, no matter how small, when there is no table there, because then the table would not have spatiotemporal continuity.

Secondly, the idea that "pushing a button on the machine will cause any table at that location to vanish into thin air" needs explanation in physical terms. If it was just magic, we could say that if magic occurs, it falsifies the theory - but magic does not occur; therefore, the theory stands.

Shoemaker says a "machine" could theoretically do this work, so what kind of machine could it be? Let us imagine that it is a machine that uses an energy beam to break down objects to their atomic components. In the case of a wooden table, the atomic components would be carbon, hydrogen and oxygen. Such a machine would certainly be very powerful. It would give the appearance that the table had disappeared into thin air, except I imagine there would be an explosion as the atomic carbons reacted with the oxygen and hydrogen. So what would really need to happen, would be for those atoms to be fired off at high speed away from the area. In order for the carbon atoms to not explode on contact with other atoms, they would have to be bonded with other carbon atoms into miniature diamonds or the like. It might still be a bit like an explosion. Perhaps if the carbon was joined with the oxygen in the air to form carbon dioxide, and the hydrogen was bonded with oxygen in the air to form water vapour, then the table would effectively turn into a slightly steamy area of the room. Of course, there can't be a moment when it is just a slightly steamy area of the room, because then there would be no spatiotemporal continuity with a continuing table.

There is still the problem of getting the new atoms in there to form the new table. Those atoms would have to be arriving just as the old ones were leaving. All this atomic to-ing and fro-ing is unsatisfactory because Shoemaker needs there to be spatiotemporal continuity without the existence of the same table. If there is not continuity, then we can accurately call it a different

table, because there is not continuity with the old table. Later I discuss (in the Swampman example) how, the atomic identity of a high-level object doesn't affect the high-level identity of the object. That being the case, let us have the machine do things differently: The machine breaks the atomic bonds, as before, it then uses its own pattern to put the bonds back together, using those same atoms as are already there. There is still a residual issue that there is a moment in time where all the atomic bonds in the table are broken, and this would count somewhat against its spatiotemporal continuity.

Here is another formulation: The machine identifies the area where the table is to be removed, scans the area, and then changes the atomic structure of the table as is required to create the new table. As it turns out no change is required and so nothing is changed. The table, at last, maintains spatiotemporal continuity, but is a 'new table'!

So finally we have the logical claim that it is a new table:

> If t is the time at which the buttons [on the machines] were pushed, then the nature of the table-stages that occurred after t is due to the pushing of the button on the table producer [machine] at t, and not at all due to the properties of the table that was there before t; given that the button was pushed, we would have had such a table there after t, even if there had been no table, or a very different table there before. It seems plain that in this case one table has been replaced by another. [Shoemaker, 326]

In the above example, the machine has changed nothing about the table, as that would violate its spatiotemporal continuity. But now, it is still being claimed that the table's continued existence is not due to its continuous nature, but rather to the activities of the machine. If no table had been there, the machine might have formed a table out of 'thin air' using the carbon, oxygen and hydrogen atoms in the air, but this has no impact on the existence of the same table. We must achieve the first step in the argument before the next part works. Shoemaker needs to find a way of creating a change consistent with the table having spatiotemporal continuity, without using magic. Shoemaker's solution is… A miracle occurring, as shown by Shoemaker's next example, which is the idea of God deciding that:

> …at a certain time a certain object - let it be a stone tablet - will disappear into thin air then…he issues another decree, this one to the effect that at that time there will come into existence at a certain place a stone tablet of a certain description - and by coincidence the place and description are precisely those of the existing tablet at the time of its agreed annihilation. …
> What our deity has inadvertently done is to replace one tablet with another just like it, and in such a way as to preserve spatiotemporal continuity. [Shoemaker, 326-7]

That is what Shoemaker describes the deity as doing, except there are a couple of problems.

1st. It relies on a miracle occurring, which is much more than even an extremely improbable event requires.

2nd. If the sequence of events described by Shoemaker actually did occur, and we didn't have access to the divine, then there would be no physical way to verify that anything had happened. Therefore it is not verifiable in scientific terms.

3rd. As a logical argument, the conclusion that God has replaced "one tablet with another just like it" does not follow, except by definition, from the premises of the following argument:

P1: A stone tablet disappears.

P2: A stone tablet appears.

P3: The second tablet is different to the first tablet.

P4: There is spatiotemporal continuity.

Therefore: There can be a change in identity while preserving spatiotemporal continuity.

Firstly, I would accept the claim there is spatiotemporal continuity. With P1 it is said that a stone tablet disappears, but it does not 'disappear' in any ordinary sense, because there is no way to detect any change. At the point that it 'disappears' nothing seems to change. Similarly, when the next one appears, there is no apparent change. Then we have the claim that the second tablet is different from the first. It is only supposedly different because Shoemaker said that it *is* different, which is the very claim up for discussion.

I would say briefly that we may have a concept of something, such as imaging that a person can fly by flapping their arms, but the imaginative concept does not present a challenge to physics when discussing the reality of human life.

An argument about the difficulty of presenting cause as necessary, is that if I argue that if causality is something that is uniquely required by identity, then it should be easy to show that it is so. It should not need any elaborate science fiction or theological examples to prove that cause is necessary. Most philosophers seem to take it as obviously being the case that causality is necessary, but when it comes time to prove that this is the case, then there is difficulty proving it. This seems to be a strange result for something that is supposed to be very central.

Chapter conclusion

In this chapter, we have seen that philosophers can propose either similarity or cause as fundamental to identity. We have seen that it is challenging to prove the existence of cause as being of importance in identity, although it is widely assumed to have such importance. Therefore we can see that the claim that causality is important in understanding identity is severely weakened.

Chapter 2 - Can we remove cause, or remove content?

If causality is not essential in identity, we should be able to remove causality and still have the same identity understanding as before. This chapter finds if we can achieve that by introducing three arguments. Below is the second part of the argument previously presented:

P2: We can deprive entities and objects of causal connections without changing our description of their identity.

In order to see if causality is necessary to identity, I use Russell's "five minute world" example. This example is a crucial step in removing causal history from being conceptually necessary to identity.

Argument 1: The consequences of a five minute world

P1: The world in its current form could have been created five minutes ago.

P2: *If* the world in its current form *was* created five minutes ago, we would not propose a change in the identity of things around us.

Therefore: The difference between the world being created five minutes ago, and the world having its currently assumed age is not a conceptually necessary one for the assignment of identity.

Furthermore: *If* the world being created five minutes ago, *and* having it's currently assumed age cannot be a relevant one identity wise, *then* prior causal history cannot be a relevant factor in identity.

The first premise we must examine is, **P1**: The universe in it's current form could have been created five minutes ago. Russell's text on that I will quote in length because Russell is a famous proponent of the five minute hypothesis as a sceptical hypothesis:

> In investigating memory-beliefs, there are certain points which must be borne in mind. In the first place, everything constituting a memory-belief is happening now, not in that past time to which the belief is said to refer. It is not logically necessary to the existence of a memory-belief that the event remembered should have occurred, or even that the past should have existed at all. There is no logical impossibility in the hypothesis that the world sprang into being five minutes ago, exactly as it

> then was, with a population that "remembered" a wholly unreal past. There is no logically necessary connection between events at different times; therefore nothing that is happening now or will happen in the future can disprove the hypothesis that the world began five minutes ago. Hence the occurrences which are called knowledge of the past are logically independent of the past; they are wholly analysable into present contents, which might, theoretically, be just what they are even if no past had existed. I am not suggesting that the non-existence of the past should be entertained as a serious hypothesis. Like all sceptical hypotheses, it is logically tenable, but uninteresting. All that I am doing is to use its logical tenability as a help in the analysis of what occurs when we remember.[25]

To discuss, I will go through the argument line by line:

R. P1: "…everything constituting a memory-belief is happening now, not in that past time to which the belief is said to refer."

This, I think, is an uncontroversial premise.

R. P2: "It is not logically necessary to the existence of a memory-belief that the event remembered should have occurred, or even that the past should have existed at all."

I would say; Never mind *logically* necessary it is not even *physically* necessary, that the event remembered should have occurred. However, we may get into a semantic issue if people choose to define memory in causal terms. It might be claimed, for instance, that if the event did not occur, then it is not actually a memory, but has some other epistemic classification, such as where Parfit talks about quasi-memories. [R&P 220] In terms of the necessity of the past having existed at all, this is also a disputed concept - for instance, with Davidson, in terms of the Swampman hypotheses, which we will deal with later. It may be physically, but not logically necessary that the past exists, but this is not really an issue from our point of view.

R. P3: There is no logical impossibility in the hypothesis that the world sprang into being five minutes ago, exactly as it then was, with a population that "remembered" a wholly unreal past.[26]

---

[25] Russell, Bertrand. *The Analysis of Mind*, The Floating Press, 2008. Ebook, p. 203. https://ebookcentral-proquest-com.ezproxy.otago.ac.nz/lib/otago/detail.action?docID=432353

[26] In fact Quantum Physics describes it a physically possible. For example the Boltzmann brain hypotheses.

Some might be dispute this. It certainly might be thought that it is *physically* impossible for the world to spring into existence in that way out of nothing, although there may be some quantum phenomena that could bring such a thing about.[27] Whether it is *logically* possible is a fairly low bar. If it was the whole universe that was to spring into existence we get into multiple difficulties.

The most fundamental logical difficulty is that if the universe was to come into existence from nothing we would have to say that at 'time 1' nothing existed, then at 'time 2' the universe existed as it was five minutes ago. But it is not clear how we get from nothing to something. Another potential problem is that if the universe as it is now came into existence five minutes ago, then the theory of relativity would predict that it also existed before that since time is not a constant:

As Einstein remarked:

> For us physicists, the distinction between past, present and future is only an illusion, however persistent", and these concepts really do not figure at all in special relativity.[28]

So for the universe to come into existence a few minutes ago may not make sense from a physics point of view, but for the Earth to come into existence in that way would make more sense. Even that has problems though. If the Earth suddenly popped into existence in its present orbit around the sun, this would no doubt create immense gravitational disturbances on the other inner planets. What would be better is if there was an equivalent mass in place of the Earth five minutes ago, and then it was shaped into the Earth at that time. A 'swamp-world' hypothesis if you will[29]. It would also solve the puzzle of why other intelligent life in the universe has not contacted us. - We haven't been here long enough.[30]

R. P4: "There is no logically necessary connection between events at different times; therefore, nothing that is happening now or will happen in the future can disprove the hypothesis that the world began five minutes ago."

It is difficult to prove a negative. Whether or not this claim is true or not does not matter for my purposes. Of more interest is that it could be discovered that *a* world *did* begin five minutes ago.

---

[27] Boltzmann brains are also a possibility but they suffer from some of the same problems as idealism. Much better to have the whole world pop into existence.

[28] Einstein's letter of March 1955, http://www.physicsoftheuniverse.com/topics_relativity_spacetime.html

[29] See Discussion of Donald Davidson's Swampman elsewhere in this thesis

[30] As well as a 'mass quantum event' we could imagine the situation where there was the big bang which then caused matter to randomly coalesce into the form we see currently without the normal causal processes being in play.

For instance, in terms of the simulation movie “The Matrix” - The matrix is turned on, and then five minutes later there is the situation that the apparent world began five minutes ago.
Even within the context of our own planet starting its existence five minutes ago, we could soon discover (for instance, from an alien civilisation watching us, or from our own observations) that actually the planet began its existence five minutes ago - or whatever the relevant timeframe was.

R. P5: “Hence the occurrences which are *called* knowledge of the past are logically independent of the past; they are wholly analysable into present contents, which might, theoretically, be just what they are even if no past had existed."

Above is a critical passage. It is a bit difficult to understand, but the basic point seems to be that things could be exactly the same, even if there was no actual past. But more importantly, by his emphasis on the word “called”, Russell seems to be saying that what we think is knowledge is not infact knowledge.

So where to from here? If we amend argument 1 to:

Argument 1.1: The consequences of a five minute world, updated

P1: The Earth in its current form could have been created five minutes ago.

P2: If the Earth in its current form was created five minutes ago, we would not change how we live.

Therefore:

C: The difference between the Earth being created five minutes ago, and having its currently assumed age cannot be a behaviourally relevant one.

We have been through Premise 1 of 1.1. Premise 2 is a more speculative endeavour.

If the earth in its current form was created five minutes ago, we would not change how we live.

First of all, I am sure that it would be universally agreed that if no one knew that the world had been created five minutes ago, then it would not affect how people behave. The only question is what difference it would make if they *did* know. Only humans could have the capacity to know and understand such a situation. Therefore, only humans could be affected by that knowledge. Other creatures would continue on as before.

On the face of it, a revelation that the world had only been created five minutes ago would cause a complete breakdown of society, because everything is supposedly based on prior events with a certain causal history. For instance, the laws having been passed through parliament etc. However, if everything was to be simply redefined in terms of things being causally and content-wise *as if* the earlier events had happened, then meaning could be brought to the situation and the crisis averted. So in that instance we could say: There is no past, but everything exists *as if* there is a past. Therefore; We will continue to act as if there is a past. Given that the alternative was societal breakdown, when a perfectly suitable means of avoiding the societal breakdown existed, surely we would wish to avoid societal breakdown? It is of course possible that someone could 'bite the bullet' and prefer societal breakdown to what they may regard as 'living a lie', but I do not regard this as a rational option.

Argument 2: The difference in the five minute earth is cause not content

We still need something else to link the argument together. That is, that it is actually content similarity which remains with the five minute earth, but not causal history.

P1: If the world was created five minutes ago, it would have all the same content as it does now.

P2: If the world was created five minutes ago, it would have a different causal history than it does now.

Therefore: The difference between a world created five minutes ago and its currently assumed age is its causal history.

With Argument 2, Premise 1, I see no reason to challenge that the content would be any different. With Argument 2, Premise 2; If the world was created five minutes ago, it would have a different history i.e. it would lack the current history.

Content similarity over time - Sealing off one possible objection.

One possible objection[31] to a five minute world would be that if a dispute arose about what happened in the faux past, then it would be in essence unresolvable. There would only be the past in terms of current evidence about the past - in terms of information in the world now. That is all *we* have to access the past. But someone might imagine using a time machine to go back and check on a historical event. Since there would be no time to go back to, there is at least one consequence of the five minute world premise: Somewhat less enjoyment of the *Back to the Future* movie series. This might be presented as a counterfactual of the counterfactual; That being, that one loses something from being in a five minute world - the conceptual possibility of time travel back into the past.

Argument 3 - The five minute world and the possibility of time travel

P1: If the Earth was five minutes old, time travel to an earlier age of the Earth would not be even logically possible.

P2: The logical possibility of travel to an earlier age of the Earth exists in the current Earth.

Therefore: We lose at least one thing in the present by the Earth being only five minutes old.

This is a problem because what we want is for the five minute Earth to be functionally the same as the current Earth - for all the current things to make sense. It makes sense to be an historian because all the same resources are still there. It makes sense to be an archaeologist because the Earth can still yield its 'secrets' as before. But alas, it does not make sense to be working on time travel back into the past, because there is nothing there. - No Earth at least. Someone might use this to argue that we are worse off in the five minute world, and therefore content is not all that matters. This can be solved by having *content similarity over time*, in terms of any number of continuous five minute worlds, causally unconnected and popping in and out of existence, with a tiny gap between them. So even if you were to go back in time into the gap, it would be an unnoticeably short period of time until another universe came into existence to take its place. So if someone is to go back in time they find themselves in one of these worlds and content is preserved in the current world.

However, someone might say: What is the meaning of this? It is bad enough to accept *one* five minute world as a thought experiment, but how can we accept multiple, causally unconnected five minute worlds popping in and out of existence? How likely is that? I would say that it is somewhat

[31] Another possible objection that someone could give is that causal history is still of the most importance, as the content similarity is only in terms of the causal history. i.e. if we ask: Content similarity with *what*? It is with what *would* have happened if there *had* been the causal history; therefore, the causal history is still the most important, even in its absence.

unlikely, but if we see time travel as a possibility and think we may be missing out on the fruits of it in a five minute world scenario, and thus see the five minute world as a poorer and more restricted world because only content - not cause, is taken account of, then this solves that scenario. Since time-travel back into the past is potentially even less likely than multiple, unconnected five minute worlds, it should not matter about the likelihood.

Someone could object that if we knew the world to be a five minute world - even with causally unconnected content similarity (which could be known as “content continuity”). It would create problems and differences in how we see things. Yet ultimately, it is not the lack of cause that creates the problem but *knowing* about the lack of cause. Removing causal history does not change anything in how we act, and how we think about things, but as soon as we know about the lack of cause, then that seems like an issue to people if they have already decided that cause is important. If we did not think cause was important, it would not seem important.

The five minute world scenario works to show us that causality is not naturally important to us. The only thing that would make us act differently if the five minute world was the real world, is if we knew about it, but that is just valuing cause regardless of its actual effect. The only objection that can get traction is that time travel back into the past becomes impossible in the five minute world scenario, but this can be resolved by having a number of causally unconnected five minute worlds in series, so that in fact, the same kind of historical content does exist in time.

Can we remove content?

It is now time to examine whether content can be removed without identity being affected. To do this, it requires that we are able to remove all content and have something regarded as being the same thing. Restating premise three from chapter 1:

P3: We can not deprive entities and objects of content similarity without changing our description of their identity.

It is difficult to think of examples whereby something could have content similarity totally removed and still be considered as the being the same thing. In a partial sense it is easier, as on the face of it, we can demonstrably change the content similarity of entities and objects without changing their basic identity. For instance, we can take a coin and scratch it and still have it accepted as a coin.

Throughout their life, a person changes in their similarity to at previous times, but they typically go by the same name. If we have a person named “Ben” for instance, when Ben is a baby he is called “Ben”, and when he is old he is also called “Ben”. Yet every known change in content

changes the overall description of the entity. What was previously an unmarked coin, is now a coin with a scratch. What was previously young Ben, is now old Ben.

Component parts assembled into the same design

When the same component parts are assembled into the same design, we typically say that it is the same object. But we may not.

Let's say that Divesh gets a lego set and builds a truck, "Divesh's truck". Then the truck is taken apart and Norman comes along. Norman then gets the same lego set and builds the same truck, "Norman's truck". Is it the same truck? Someone might argue that although identical, the trucks are different, and their difference is a causal one. Divesh *caused* one truck to be built, and Norman caused the next truck to be built. But if we did not wish to include cause, then it would also be possible to say that the trucks existed at different spatiotemporal points and that is the difference between them.

So how do we get identity in this case: Through cause or spatiotemporally? The question then would be whether it would be possible for the truck to have the same identity without cause, or without it existing spatiotemporally. It certainly would be difficult for the truck to have the same identity without existing spatiotemporally.

What if when Norman received the truck set; he started to construct it, and an alien force possessed his arms and the lego pieces, so that as he moved to make the truck, his arms and hands were moving in parallel with his thoughts and the pieces were moving around without him actually touching them (although he thought he was)? So at the end of the process, the truck was built exactly as before, but without any actual causal link between Norman and its construction. In that case we would have the same result, but without the causal link, so it seems we do not really need it.

What about if we were to have the situation where, when he is given the lego set, Norman does not actually build the truck, but only imagines that he has built it? What if then everyone has exactly the same thoughts about the imaginary truck as they would have had about the real one? Could we then say that the truck has the same identity without existing spatiotemporally?

Let us imagine, for instance, that Norman goes home and his mother says: "What did you do?". Norman says, "I built a truck", and everyone has the same thoughts as they would have had in the first version of the example but without any apparent content that they refer to. What is the identity of the truck? Here, it seems like the identity reduces down to content similarity of description. Without cause or spatiotemporal substance, all we have left as its identity is its description. Identical identity requires the content of identical descriptive thoughts.

Part 1 conclusion

Part one of this thesis is on what identity is. I have examined various scenarios in order to see what is essential to identity, to find what identity consists of. I have found that we can remove cause from the world and still have the parts of the world retain their identity, but we cannot remove content and still retain identity. Therefore, only content similarity is necessary for identity, not causal history.

Part 2 - What matters?

Once we have worked out what identity consists of, we need to know what matters for us regarding future states. Is it the preservation of identity or something else? My answer is that it is content similarity, but not content similarity as the (content similarity) closest continuing thing on the spatiotemporal path, as is identity, but rather an area of content similarity anywhere in the spatiotemporal dimension, where the same network identity still holds. Furthermore, this content similarity should be as 'wide' as possible. I call this 'point to point' content similarity. Where we have the situation of seeking a 'further fact' about what matters after that, we are most likely to find it outside the range of narrow content, and in the content identifiers used to identify it out in the wide content world, i.e. its network identity.

While there may be a 'further fact' which helps determine intuitive identity, there is no 'final fact' which determines absolute identity in two non-identical situations. However, a further fact may be a relative fact about reference to the object in question.

Why use the term 'what matters'?

When saying "what matters" in this thesis, I am referring to 'what matters in survival'.

Using various examples, Parfit prises what matters in survival away from strict identity relation in terms of cause. One example is an imaginary teleporter which destroys someone's body and recreates a replica elsewhere. This is different from another possible imagined teleporter that would break apart someone at an atomic level and beam those atoms elsewhere for reassembly. Under a Parfitian framework, these two cases would be seen as equivalent in the result they achieved.

If the way a teleporter worked was to scan and send only the information, while the original person was left, then under Parfit's philosophy, that would be just as good as if the person travelled in the regular way. If, as under Parfit's example, someone is left alive for a short period of time after they have been scanned, then this is called 'being on the branchline'. The replicated person

is on the ‘mainline’, and they are the one to which concern need be directed. The person on the branchline has no rational reason to feel concerned. Any such fear is like a fear of heights or some other irrational fear. Believing that there are legitimate reasons for concern in such a scenario is called the *further fact* view.

I am attempting to prove that causal connection is not important in survival but that there is a further fact, and that is identity within the network. So my first task is to prove that causal connection does not matter in survival but that social or network connection does.

So first up is proving that causal connection does not matter.

Chapter 3 - The proposal that content similarity is what matters

In order to see that content similarity is what matters and not causal connection, we have to bear in mind the lesson of the five minute world, which is that it would not affect our behaviour if the world turned out to be a five minute world. The example of Swampman is an instance of a ‘five minute person’ with a random cause. If we accept Swampman as being the person, or as good as the person, then it would show that what we care about is content and not cause. I argue that whether or not Swampman *is* the person, they are *as good* as the person.

Example: Randomly created content - Swampman

Swampman is an example created by Donald Davidson in order to advance a particular thesis about meaning. I am using the example to talk about identity, and what matters, since in the example he makes claims about identity as well as meaning. The example is described like this:

> Suppose lightning strikes a dead tree in a swamp; I am standing nearby. My body is reduced to its elements, while entirely by coincidence (and out of different molecules) the tree is turned into my physical replica. My replica, The Swampman, moves exactly as I did; according to its nature it departs the swamp, encounters and seems to recognize my friends, and appears to return their greetings in English. It moves into my house and seems to write articles on radical interpretation. No one can tell the difference.[32]

---

[32] Davidson, Donald. *Knowing One's Own Mind*, Proceedings and Addresses of the American Philosophical Association, Vol. 60, No. 3 (Jan., 1987), pp. 441-458, American Philosophical Association, http://www.jstor.org/stable/3131782

Swampman is a very important example in understanding the nature of identity, and yet it has not generally been used as such. This is not to say that the same kind of issue has not been considered separate from the swampman example. For example Nozick says:

> As you die, a very improbable random event occurs elsewhere in the universe: Molecules come together precisely in the configuration of your brain and a very similar (but healthier) body, exhibiting complete psychological similarity to you.
>
> This is not you; though it resembles you, by hypothesis it does not arise out of you. It is not any continuer of you. In earlier cases, by psychological continuity I meant "stemming from" and "similar to" of course, we can have the first without the second, as when drastic changes in psychology are brought on by physical injury or emotional trauma; case 7 shows the second without the first.[33]

So Swampman is like the Nozick example, except there is not the huge spatial gap between ourselves and the randomly cohering being in the Davidson example. I suspect that this gap is in order to create further intuitive distance between ourselves and the randomly cohering being. Davidson attempts (bizarrely, in my opinion) to create this kind of intuitive distance by saying about Swampman's creation it is "out of different molecules" as if specific molecules could be the 'carrier' of identity and in this case, meaning.

I see the Swampman example as giving one of the clearest examples of why it is content similarity that carries identity, and what matters, and not causal history that creates it, but it was intended as an example to illustrate the exact opposite. I don't blame Davidson for this, in fact I celebrate it. It reminds me of another famous case regarding Einstein and quantum entanglement related by way of (lack of) causality:

> "The notion that randomness, indeed absence of causality, lies at the heart of things is unsettling in the extreme, and there's some comfort in seeing that Einstein shared our instinctive reluctance to countenance it."[34]

> "If Einstein isn't always afforded due praise, however, that is at least understandable. For he 'discovered' entanglement via a thought experiment, that because it posed an apparent paradox, demonstrated in his view that such behaviour couldn't possibly be real. Einstein wanted to bury entanglement even as he unveiled it." [Ball, *Beyond Weird.* p162]

Swampman, meaning and value

---

[33] Nozick, Robert. *Philosophical Explanations*. Cambridge, Mass: Harvard University Press, 1981, p.41.

[34] Ball, Philip. *Beyond Weird.* The Bodley Head, London, 2018, p.161.

Davidson says about the creation of a ‘Swampman' replica:

> My replica can't recognize my friends; it can't recognize anything, since it never cognized anything in the first place. It can't know my friends' names (though of course it seems to); it can't remember my house. It can't mean what I do by the word ‘house’, for example, since the sound ‘house’ Swampman makes was not learned in a context that would give it the right meaning—or any meaning at all. Indeed, I don't see how my replica can be said to mean anything by the sounds it makes, nor to have any thoughts.[35]

Swampman is an argument against content similarity as a main criterion of value, where value is taken to be meaning. The reason for this is that we have content similarity, but there is proposed to be no meaning and presumably no value.

The Swampman argument is false and it is easy for me to see why, but the problem is that where the value of meaning is in its meaning something to somebody, if someone simply denies the meaning they can maintain the argument.[36] The argument is supposed to work by saying that since meaning is often thought to be derived by causal connection, once that connection is broken there is no meaning able to be derived or generated. Even if there is apparently ‘meaning’.

I take it to be a natural and uncontroversial extension of the Swampman argument regarding meaning, that the swampman is not seen objectively to have the same identity as the original person, regardless of whether he *appears* to have that identity. This obviously goes for having what matters about them also. I maintain that the swampman may not have the same identity as per definition in the example: “My body is reduced to its elements” but that the swampman has *what matters* about the person. It is, in Parfitian terms “as good as ordinary survival”.

Supposed counterexamples to Swampman’s identical rights

Let us take a look at some examples where being identical alone does not seem to be enough to maintain all of what matters about a person.

For instance, let us imagine something like the famous Twin Earth example but less complicated: On Earth we have Jim, who when he was born was born in what is called a “line house” on the

---

[35] Davidson, Donald. *Subjective, Intersubjective, Objective: Philosophical Essays Volume 3*. Oxford: Oxford University Press, 2001. Print ISBN-13: 9780198237532, Published to Oxford Scholarship Online: November 2003, DOI: 10.1093/0198237537.001.0001

[36] There is another aspect to this and that is that it adds something to Descartes meditation on first philosophy in that if Davidson is correct we can know if we are Swampmen right now. Just think “I think therefore I am". If you can achieve that then you’re not a Swampman, because a Swampman would, by Davidson’s understanding not be able to think it. Therefore, by that rationale there is something else that we can fundamentally know other than that there is a thinker - that the thinker has not randomly coalesced.

American/Canadian border. Jim's mother gave birth on the American side of the border. Jim on twin Earth is currently physically identical to Jim on Earth, except when he was born his mother gave birth on the other side of the house, on the Canadian border side. The Jim's don't know this however, they both presume they were born on the American side of the border and are American citizens. The fact remains that though they are physically identical and their worlds are currently physically identical (By some kind of confluence of physical events), when one says that they are American one is telling the truth and the other telling a falsehood. Although currently identical, a different causal history accords each one of them different rights. One brother can legally run for president and one can not, for instance.

It then stands to reason that if there is an identical swampman replica of Jim, and separately an identical swampman replica of twin Earth Jim, that the swamp man could not have any citizenship at all, because being identical alone could not mean anything in terms of citizenship rights.

Here's another example: Let's say that there is a murder, and the police are looking for the murder weapon which is the crucial evidence in the case. They recover the weapon they are looking for, but on the way back to the station, with said weapon, the police car crashes and catches fire so that the weapon is destroyed. Could the police then go down to a local gun shop, purchase an identical firearm and present that to the jury, say it is the murder weapon and get a conviction? We would surely regard that as gross police misconduct. So it looks like we have a class of things, which are true of those things alone and not of objects that are physically identical to them. It looks like content similarity as a sole criterion of value is in trouble. But is it? There are two ways to solve these problems. One is by reasonableness with regard to the narrow content, and the other is by wide content.

Reasonableness and how selection creates meaning

The concept of 'selection', as described below, creates a mechanism whereby randomly created content can still have the same meaning as causally created content. That mechanism is selection on the basis that the content is the same as causally created content.

In the realm of computers, it is typical to take a copy of something to have all the veracity of the original. Nothing is gained from having the original. In practical terms, content similarity seems to be all when it comes to computer files.

Normally, we agree that when a text is copied, the copy takes on the causal history of the original. If we did not think that, then plagiarism would be acceptable. Under that scenario, someone could plagiarise, and the act of copying would sever the causal history of the copy from the original.

In other cases of manufacture, there may be no 'original' as such. I may hold up a first edition of a book and say "I don't want this copy of the book, I want the 'original'. Not the manuscript, but the

original book". In that case, there is no original. There is the first book to be manufactured, but that is seen as irrelevant, and it does not typically have any greater value than the rest of the books manufactured at that time. So wanting the 'original' of a book is just unreasonable.

What about a 'swampbook'? In this case, we don't have to speak of a book found in a swamp, because of the use of computers. Imagine that lightning hits a computer, and there is a power surge that causes a corrupted document file. In this case, the corrupted file happens to be identical in content to an already published ebook. In that case, the 'swamp ebook' would not have the causal history of an ordinary ebook, and so under the Davidson schema it could not 'mean anything' or have any of the copyright implications of the ordinary ebook. This could provide an opportunity for something called "swampware".

Swampware would be a kind of software that randomly generates computer files with random characters and spaces in them, and then compares them against existing files. When it finds a match, then bingo! The user has what they want in terms of the value of the text. We can imagine a computer that works incredibly quickly, so a program is run and it starts outputting swamp ebooks, webpages, images, and computer code: All copyright free! This seems to present a major issue. It seems like it would be difficult at this point to maintain the intuition that the swamptext has no meaning. It just seems like a convenient way to get past the copyright laws. What is it about the swampware program that allows it to create useful documents? It seems like the process that creates it is the selection of randomly generated text against the original. But if swamptext has no meaning to begin with, how can selection assign it meaning? Well, just as other identical things can have facts about them, so can swamptext have a fact assigned to it. That fact is that it is identical to text with a causal history and should be treated as such - not to do so would be absurd. People can deny that if they wish. If they deny it it suggests they would permit the generation of their works by means of a swampware program.

To further extend the example, let us imagine that it is not text that we are creating, but human beings. A machine is quickly creating swampmen and comparing them against the original. It is then destroying them unless they match. We then have the same situation with the swampmen as with the swamptext. But this swampman is in the same situation as the original Davidson swampman. The only difference is that with this swampman he is retained by virtue of similarity with Davidson, as opposed to just being created and happening to be identical. It is not a significant difference. Of course with the original swampman, the only reason that his friends and acquaintances take him seriously is that he looks and acts as Davidson. So that is a kind of selection. If it was just some monster crawling out of the swamp then there would be no issue to begin with. So for Davidson to maintain that the swampman has no meaning, he would have to maintain that swamptext has no meaning and anything that is randomly created has no meaning, and that has real-world effects as we have seen.

If we go back to the citizenship case where two physically identical people have different citizenships: If there is a swampman replica of either person, the citizenship would have to be indeterminate. But that is only because the criterion of citizenship is unreasonable to begin with. This is just like how two Australian senators resigned because of the unreasonable citizenship criterion that applied to them in the Australian constitution.[37]

In the case of the murder weapon example it would be completely reasonable to provide a *perfect* copy of the murder weapon - either a swamp copy or a copy via causal means, if the ambiguity in the original example was removed. For instance, if the police could attend a crime scene and take an atomic level scan of the evidence at the scene, then if the original evidence was discarded or destroyed the copy would be good evidence. This is just as crime scene photos are good evidence - they are not the original crime scene.

There is a famous murder case in New Zealand where a lot of the evidence centred around the rifle used for the killings. Imagine that the police did an atomic level scan of the Bain[38] crime scene, and they were able to draw down that information and produce an atomic replica of the rifle used in the Bain family murders on demand for court cases. In that case the replica would be of more value and use than the original rifle, which may lose some of its value as evidence after repeated handling and the natural degradation that comes with age, over time.

We have seen that when we select a randomly generated replica of a person or thing against the original, it is as good as the original or a causally created copy. It is just another means of doing the same thing, and it is only reasonable to see copies in the same light as their originals so long as they are exact copies.

Random signs and selections

Davidson says: “I don't see how my replica can be said to mean anything by the sounds it makes”. This is a very important point, and the question is: Can randomly created messages mean anything? I say they can, where their content and context are the same as a ‘causally’ created message, or via selection.

Positive selection

---

[37] Karp, Paul. *Australia citizenship crisis reignites as senator and four MPs quit*, The Guardian, 9 May 2018. https://www.theguardian.com/australia-news/2018/may/09/australia-citizenship-crisis-reignites-as-senator-and-four-mps-quit

[38] The Bain murders were a famous New Zealand crime where there was a problem determining whether the son or the father did the crime of killing the rest of the family.

Imagine this scenario: There is a metal panel with plastic letters with magnets on them that can be arranged in any particular way. If a young (pre-literate) child of two plays with the letters and makes a word, or even a sentence, then we would not say that they mean it. The child cannot mean anything by what he or she might 'write' on the board by playing with it. If the child writes a swear word, for instance, we should not get angry with them. The child is in the same position with regards to meaning that Davidson says swampman is. However, that does not mean that the words the child 'writes' cannot have any meaning per se. If the child randomly writes a swear word and we take a photo of it and post it on someone's door, then that has meaning. We have given it meaning by our actions in selection. In this case, a positive action creates the selection.

Negative selection

Consider another case: A lift repair technician has a sign in his bag which says "Danger! lift out of order." In this example, the lift repair technician has a dental appointment after work with his son in a high rise building. While he is waiting for the appointment in the building, his young son takes his sign out of his bag and places it in front of one of the lifts. The sign says, "Lift out of order" The father sees this and takes no action, so people start to avoid using the lift.

Is the off duty lift technician in any way liable for this inconvenience? I believe the answer is "yes" because by leaving up the randomly placed sign, he is sending the message that the lift is not functional when he believes that it is. So this would form a kind of negative selection. Where a randomly placed sign is left in place, although no action is taken, the lack of action is the action.

With Swampman, meaning is even more fundamental. It is not just that we give the utterances of Swampman meaning by selection, they have meaning themselves by virtue of the initial conditions of the example as Davidson constructs it - that being the output of the brain of a human duplicate.

So if a lift technician is standing outside a lift, warning people not to use it, and he is hit by lightning and replaced by an identical swampman doing the same thing, then we should treat the warnings of the swamp technician with due regard. It is incredible to me that anyone could say otherwise. It is the very fact that he is identical that is the issue. It is definitional. If another identical thing replaces anything with meaning, then that identical thing has the same meaning by virtue of its identicality.

We can set the argument up like so:

Premise 1: A lift technician is standing outside a lift, warning people whether the lift is safe to use.

Premise 2: The lift technician is replaced by an identical 'swamp technician'.

Premise 3: We know the above.

Proposition 1: If we give an utterance serious regard, then it has meaning.

Proposition 2: We would give the warnings of the swamp technician due regard.

Therefore; The swamp technician means something by the sounds he makes.

Importantly, it is not just because we had selected and created meaning like in the child's letter pattern situation, which always had meaning to begin with, by means of its identical nature.

It could be that none of this matters to some people - that they are such a follower of Davidson that they are willing to step into the lift to their doom, regardless of the desperate pleas of the swamp-technician. If this is the case, there are still further problems that await such a person.

Non-causal infection

If we have a swampman go out into the world and partake in normal activities of work and leisure, then, if we are to take cause seriously as the only source of meaning and value, we have a problem that arises from one randomly created being not having the right kind of causal history as part of the causal chain in general. The argument is simply that if one part of the causal chain does not have the right causal history to be legitimate, then none of the resultant parts of the chain are legitimate. For instance, if there is a swampman whose job was working for a bank and he had to make decisions about who would be lent money, then from the causalist's point of view, those transactions would be equivalent to a bank error, and the bank would be justified in asking for the money back. If someone went to a furniture store and selected a bed from a swamp-salesperson who said that he would help them take it out to the car and then they could go back inside and pay for it. That would be shoplifting because they had taken the bed out of the store without permission and without paying for it. If the apparent head of state was a swamp-person then none of the laws would be legitimate because they would not have been signed by the head of state. These are just some auxiliary problems with having a swamp-person in a causal chain from the causalist's perspective.

Wide content

The second way we can address the issue with Swampman is by using wide content. Wide content means the content that is beyond just the individual.

If we have Jim, who is an American citizen by birth and a newly created duplicate of Jim, then it might be reasonable to say that Jim is an American citizen, whereas Jim's duplicate does not have any citizenship at all - unless by 'birth' we mean wherever Jim's duplicate was created, which might be America or even such an unlikely country as New Zealand.

In any event, Jim's duplicate cannot duplicate the citizenship of Jim, by how we normally decide citizenship. It seems like it might be unreasonable to decide citizenship in that way, but that is how it is. Regardless of that however, a causal link is not actually required in order for Jim's duplicate to be seen as a citizen - just wider content duplication with Jim, so that where people currently see the original Jim as a citizen, they must see the duplicate Jim as a citizen.

There is an external knowledge criterion, so that when the network of people outside of Jim and Jim's duplicate, know that one of them is a swampman, and one is not, then they can make a discrimination, not on their current properties but on the extrinsic properties that they themselves apply. Thus it is the social identity or the network's identity of Jim that is at issue.

It is the introduction of the concept of the swampman Jim that is a difference in the wide content beyond Jim. Where this is absent, then swamp-Jim will be regarded as a citizen regardless of his swampman status.

At that point, it might be thought, that although people would see Jim's duplicate as a citizen, he still would not actually be a citizen. I would remind people that citizenship is a social property and not a natural property. If the only difference between two entities is the extrinsic property of the name they are given, then that is a problematic element in identity. It would mean that renaming something gives it a new identity in a significant way. In the case of Swampman, the name "Swampman" is seen as unusually significant.

The usual way to determine citizenship, or identity in general, is the entity being the closest continuer of the person who has the citizenship assignment. In the case of the swampman, it is a closest continuer of the original person. We have no specific definition of how fine-grained a continuer needs to be.

Swampman - Explaining the controversy

The core problem with Davidson's example is that he himself defines the selection by saying it is identical. If he had just said 'something in a swamp that looks and acts like him' then there would not have been such an issue. It is the very fact that he baked the identicalness into the example that is the problem (for him). So given the massive problems that I have sketched with maintaining a causal bias in understanding and explaining swamp creatures, the question for me becomes; not how this problem is dealt with, but why people do not automatically accept that identical swamp-people are just like us and have the same kinds of rights and responsibilities? I have reviewed the relatively extensive philosophical literature on Swampman, and the reason I think that they have

not generally done this is that the causal paradigm is so massively explanatory of the physical world that to throw it away based on a thought experiment is unpalatable to most people.[39]

The literature seems to break down into three kinds of responses:

1. Denying Swampman's humanity e.g Davidson.

2. Saying the example is implausible and not worth thinking about e.g Daniel Dennett "Swamp-man is not logically impossible, just not worth discussing." [40]

3. Accepting Swampman's humanity. Such as Louise Antony.

Some people have taken the route from 1 to 2, such as when pressed on whether eating Swampman for meat was wrong David Papineau said "Merely Possible Cases are Irrelevant"[41] It is amusing to quote the anecdote:

> I was originally roused from my slumbers by a graduate student at King's College London, Eilert Sundt-Ohlsen. He was unhappy with the standard teleosemantic dismissal of the Swampman intuitions. Pressing the point, he challenged me about eating Swamppeople. He argued that, if they have no mentality, as teleosemantics implies, then it would seem to follow, absurdly, that it would be all right to kill Swamppeople and eat them as meat. This objection stopped me in my tracks. It is one thing to argue in the abstract that a good theory of representation should be allowed to override everyday intuitions about Swampman's mentality. But when we are forced to consider the ethical consequences of this decision, as I was by Sundt-Ohlsen's question, then we seem to end up with the wrong answer. If we did come across a Swampman, it would clearly be wrong to kill it for meat.

His answer to this conundrum is to change to the 'it's irrelevant" school of thought.

> I take the central core of teleo-semantics to be the claim that the belief and desire roles are realised by selectional states in the actual world. This claim is perfectly compatible with the idea that those roles might be differently realised in other possible worlds, and that in those worlds we would then care about something other than selectional states.

---

[39] It is a similar situation in quantum mechanics.

[40] Dennett, Daniel, "Cow-sharks, Magnets, and Swampman", *Mind & Language*, March 1996, Vol.11(1), pp. 76-77

[41] Papineau, D. "The Status of Teleosemantics, or How to Stop Worrying About Swampman." *Australasian Journal of Philosophy*. 79 (2001): 279-289.

I have to say that at least that is the moral response. If there is a creature that is in every way the same as you, but of different origin, then despite it being inconvenient for your semantic theory, you will let it live. A problem I see with that response, is that it is a theory which is contingent on whether or not some unlikely event actually occurs. This might be called the 'Platypus theory'; What if there was a mammal that laid eggs like a reptile and had a bill like a bird? 'Don't worry about it - such a creature is only a possibility, and we will worry about it when we see one.'

I have already argued that swampman should be taken seriously if he exists. However, the issue with 'merely possible cases are irrelevant' is that it is equivalent to saying that counterfactuals are irrelevant. Counterfactuals cannot be irrelevant. If they were, it would make health and safety, and disaster planning irrelevant, as there would be no need to concern yourself with what *could* happen.

A much better catchphrase would be 'merely impossible cases are irrelevant'. That is to say that if it is some supernatural intervention that is the example (such as with Shoemaker's example of God's stone tablets), then that is irrelevant because we don't think it is possible to begin with. Since swampman is thought to be physically possible, then it is relevant along with other possible scenarios.

Accepting Swamp-people

In terms of accepting Swampman into the human family there are two papers that I know of. One of them is Louise Antony's Equal Rights for Swamp-persons.[42] Here she makes the statement that:

> I think it's appalling the kind of narrow-minded (or should I say wide-minded?) prejudice Swamp-persons face today… Stop stigmatizing spontaneous coalescence, that's what I say. [Antony, 1996. p. 70]

After stating:

> I have no doubt that some biological categories, like species categories, are de facto historical, and I therefore concede that Swamp-persons are not, *strictu dictu*, human beings. [Antony, 1996. p. 72]

She says:

---

[42] Antony, Louise. "Equal Rights for Swamp-Persons." *Mind & Language*. 11.1 (1996): 70-75. doi:10.1111/j.1468-0017.1996.tb00030.x [Hereafter "Antony", date and the page number]

> We need a psychology that under-writes a rationalist criterion of moral agency, one that warrants the extension of moral concern to any creature with cognitive capacities identical to our own, no matter how formed. [Antony, 1996. p. 74]

The problem is not so much granting moral rights, but how far these rights should go. Economic rights? The right to have qualifications recognised? Citizenship? The right to drive a motor vehicle? The personal rights the person otherwise has?

Even if we were to accept the full Swampman intuition - 'Swampman is Davidson', it only works if Davidson is vaporised just before the Swampman emerges. If Davidson remains unscathed but the swampman emerges nonetheless, then we might think we have a problem because how can the swampman be Davidson in one instance and not the other? I might be willing to accept that there would then be two Davidson's with equal rights, but I can see that some others might not be as generous.

Let us investigate this scenario: Imagine that one Tuesday at 3.00 pm, a swamp created duplicate of yourself drops by your house. You have a nice talk and a bit of a merry time, but then at length you say, "I really must be getting on", and you lead your swamp replica to the door and say, "Have a nice life". The reason you might do this is that you *had* all these experiences that you remember, whereas the swamp replica only *remembers* them. That is all well and good, except that on the news a few weeks later, you see an article that says that cosmologists have discovered that the world is periodically 'swamp created' and just pops in and out of existence with the same content as before but no causal link. This fact is not expected to change much, as it is regarded as merely an interesting cosmological fact, much like the findings of quantum physics. It is also announced that the last time this happened was 3.00 pm last Tuesday. Your swamp replica makes his way back to your house. "This changes everything" he says. "You're just as much Swampman as me. I've been living rough for quite a few days. Now give me your credit card - I've got things I need to buy. Also, I need to access your underwear drawer." About then, you realise that you also need to access your underwear drawer.

I hope to show with this rather far-fetched example that our intuitions about the matter are not actually powered by causal history - because if the causal history is taken away, we still feel like we should have more rights than our 'apparent' duplicate. What I believe is at the root of the intuition is the feeling that another person should not have any of what we individually possess now, without us having an option to avoid that outcome or receiving something in return. But how to justify it in any other terms? First of all, the original 'embedded' person is the closest continuer in content similarity, not in narrow content similarity but wide content similarity. We can call Relevant wide content similarity "Network identity", as in the network of connections and information that apply to

the narrow content of the original, and not the duplicate as it follows a certain path through space/ time.

I believe the concept of “Network identity” solves the peripheral problems around using content similarity as a means of identity. It is another way of saying that the information held in ‘wide content’ about an entity is appropriate to the entity in question.

I will have more to say about the concept of Network identity later in this thesis. So far, what we have learned is that randomly created beings can have meaning and value, by virtue of being selected as having it by other beings. Now we must move on to more about causality. That is, where philosophers have attempted to show that causality is necessary for what matters in survival and where I attempt to refute these claims.

Chapter 4 - Pro causality hypotheses refuted

Due to the incredible explanatory power that people believe causality has, any situation whereby people propose a content-similar situation occurs without causality e.g. Swampman, has seen some philosophers wish to argue against it. I am going to go through some of these arguments to refute them.
We have already seen that causality is not generative of identity. Might it be part of what matters in survival?

Desired outcomes are not illegitimated by process errors

For people who believe in causal history as a source of identity and value, for something to have the right causal history, there must be a continual chain of cause and effect, whereby the end result is a result of all the relevant causes. However, in the case of fortuitously accurate process errors this may not be the case.

Fortuitously accurate process errors are where there is a particular system, and there is an error in that system, but by coincidence, the error is the same result as was sought. An example would be if ordering in a restaurant, and a waitperson fails to take an order correctly, resulting in the order not being placed in the system. Suppose also that the chef accidentally makes an extra meal, and so the meal is then placed on the counter and delivered by the waitperson. The question is whether this coincidence of errors should impact the person who has ordered the meal. For instance, should the customer have to pay for their meal? Incredibly, there are philosophers who believe not, as the causal chain is broken.

Another example is a vending machine which is malfunctioning, so that when it receives money it does not vend the selected food, but instead randomly ‘vends’ a particular snack food every so often. Let us also imagine that someone pays the machine, and they believe they have selected a particular snack food, and then at that time, the machine randomly vends that food. The person walks away happy and unaware, even though there is no connection between their action and the food they received. The owner of the machine is also happy because they got the money. My view is that the machine malfunctioning, at that particular time in that particular way had no negatives for anyone, so there is nothing about that particular distribution of goods that needs to be corrected.

Argument against the importance of causality

P1: There could be a situation where there is an error, whereby the correct chain of cause and effect is not followed.

P2: We believe that if the desired outcome is generated, it is not illegitimated by it not having the right causal basis.

Therefore: Having the right cause is not necessary for a legitimate outcome.

I think it is clear that people do not generally care *why* things happen, so long as they are the desired outcomes. This suggests that in such cases as the swampman case, the semantic outputs of a swampman would be accepted as legitimate if they are the outputs that are sought.

Psychological Sequentialism

There is a further argument which can be made, (and in fact has been made) by Anthony Brueckner and Christoper T. Buford[43] in response to Scott Campbell, who argued for a view he calls “psychological sequentialism”:

> This is the view that a causal connection is not necessary for what matters in survival. [Campbell, 2005]

Psychological sequentialism, is the thesis that cause is not necessary for survival. Interestingly, this follows from Parfit’s teleportation argument and has no reference to Swampman, even though Swampman would be the purest example of type.

In “Is Causation Necessary for What Matters in Survival?”[44] Campbell argues that Parfit has not been explicit in ruling out the lack of need for a causal link when considering the various cases of what matters for survival. In terms of psychological connectedness, he says that what is instead needed is quasi-connectedness - which is a concept very similar to what I call content similarity:

> B at t2 is quasi-connected to A at t1 if and only if, while B is not psycho-logically connected to (or continuous with) A , B ’s psychological states at t 2 are just as they would be if B was psychologically connected to A . [Campbell, 2005. p. 381]

---

[43] Brueckner, Anthony, and Christoper T. Buford. "Against Psychological Sequentialism." *Analysis*. 73.1 (2013): 96-101. https://doi-org.ezproxy.otago.ac.nz/10.1093/analys/ans138

[44] Campbell, Scott. "Is Causation Necessary for What Matters in Survival?" *Philosophical Studies*. 126.3 (2005): 375-396. P. 375.

He specifically lays out a scenario whereby there could be an error in the transportation machine, but that the error (by some fluke) actually produces the correct outcome.

The full example:

> Suppose that you step into the teleporter, and your body is scanned and destroyed. The information about your body state is sent to the receiver, and stored in a memory chunk. Suppose that in the next memory chunk there is a randomly generated set of properties set aside for use the next time someone presses the ‘‘Produce Random Person’’ button, and by an incredible coincidence of more-than-astronomical proportions, this information exactly resembles the information about your body. Once the receiver has received and stored the information about you, it goes into production mode, and a signal is sent to memory telling it to send the recently-received information about your body to the production unit. But suppose that a tiny short circuit occurs at this point, and the request is directed to the next memory chunk, which contains the ‘‘Random Person’’ information, and this information is sent to the production unit instead of the information about your old body. Because of this tiny glitch, ‘‘Random’’ is produced instead of Replica. But Random is exactly the same as Replica would have been, had the glitch not occurred. Unlike Replica, Random is not psychologically continuous with you. But Random is psychologically sequential with you. [Campbell, 2005. p. 381-2]

To summarise: Information about a random person is stored in the machine and accidentally retrieved instead of the correct information. The two sets of information are identical. This is a good example, because for someone to oppose it, they would have to suggest that there was a reason to rerun the machine and follow the ‘correct’ process if the error was to occur. There could be no such reason. In response to this line of thinking, Brueckner & Buford give two examples of why cause is supposed to matter. The first one is quite complex. First of all, they give the example of workers on a production line making machines, “H-gears” and “G-gears”. Fred makes a defective H-gear, but it is accidentally knocked off the production line by Wilma, who makes a G-gear (also defective) that exactly resembles Fred’s defective H-gear. When Wilma’s defective work reaches the end of the production line the boss takes it to be Fred’s work and is about to fire him until Fred explains that it is not his work. Brueckner & Buford claim that:

> This is so even though the outcome (defective gear reaching the end of the line) is exactly similar to the outcome that would have eventuated had the tiny glitch not occurred. [Brueckner & Buford, 2013. p. 100]

All this example shows is that he should fire both. Fred should still be fired because he had defective gear on the production line, that would have reached the end of the line if it had not been

for that accident, and Wilma should be fired because she had defective gear on the production line that *did* reach the end. The example does not work for sequentialism at the micro-level anyway.

Such an example does not work at the micro-level because sequentialism allows identity to be exactly replicated without cause. Therefore, objects can be associated with their creator without cause. With the H and G gears, the critical issue is who made each object, and therefore who is responsible for it. With sequentialism as the closest continuer (content similarity-wise) the responsibility is always preserved.

Of course, outcomes can be similar, with different levels of responsibility for that outcome. There is not a causal difference, but rather a difference of identity, with identity defined as the closest continuer in space/time. This is not what matters, but what matters usually follows identity except in some special cases.

What is more interesting is their second example which is the vending machine example:

> Suppose that Barney puts a dollar in a soda machine and pushes the 'Mountain Dew' button. Due to a tiny glitch, no soda emerges. Barney turns away, and a few seconds later, another tiny glitch occurs…. This time the glitch happens to lead to the release of a 'Mountain Dew'. Betty works right by the glitchy machine and is aware of its habits… Betty observes the release of a 'Mountain Dew' and grabs it. Barney, from across the room sees Betty, and he exclaims, 'Wait a minute! That's my 'Mountain Dew'!' Betty correctly retorts, 'No, it's not'… [Brueckner & Buford, 2013. p. 100]

This example needs no philosophical analysis - only a legal one. When Barney puts a dollar in the machine, the machine (or it's owners) owe him a soda. They *never* owe Betty a soda. Betty used the opportunity presented by the malfunctioning machine to steal a soda from the machine. It is then claimed that the lack of a causal link "allows Betty to fairly claim the soda that Barney wanted." She is never able to do that. Even if the machine is spitting out sodas at random with no one around, she never has any legitimate claim to them.

The argument from psychological ephemeralism

Huiyuhl Yi, in Against Psychological Sequentialism,[45] sets out an interesting example of why we should reject Psychological Sequentialism. He says:

> …imagine a countless number of random object generators spread out over the universe. The function of a random object generator is to arrange a collection of particles in a random fashion, thereby generating a material object only to destroy that object instantly. Each machine repeats this procedure over and over again. …if there really are a great number of those machines, then there

[45] Huiyuhl Yi, Against Psychological Sequentialism, Axiomathes, 2014, Volume 24, Number 2, P. 247. DOI: 10.1007/s10516-013-9221-8

> may exist a being somewhere in the universe a moment from now whose psychological state will be exactly like your psychological state a moment from now, and therefore sequential with your psychological state now…
> …Psychological Sequentialism entails that so long as you are aware of the existence of those machines, you ought not to be afraid of your death. [Yi, 2014. p. 261]

In reality, it is only their location that is the problem, as it separates them from their network identity in terms of where others expect to find them. If these ephemeral beings were simply materialised in the appropriate place and time, and stuck around, there would be no problem at all. Therefore the example utterly fails. The one thing that is true, is that "as long as you are aware of the existence of those machines, you ought not to be afraid of your death." You may well be afraid of a 'narrow content', 'fate worse than death' that the machines would create though.

Yi also has an argument about whether someone should be held accountable for doing a crime if they merely have false (non causally linked) memories of doing it, as opposed to causally linked memories of doing it. I think this is best dealt with as an externalist objection as below.

The externalist objection

Campbell talks about the Swampman case and the externalist objection to it. He refers to Swampman without mentioning (or even referencing) Davidson. Campbell's view is most like mine in that it is explicitly about what matters in survival without causality. However, his theory of psychological sequentialism leaves his theory exposed to problems that any narrow content view is exposed to. Before examining those issues, let's get an overview of his Swampman argument:

> "Suppose that at t1 you are instantly destroyed. Suppose also that on another planet, a bolt of lightning hits a swamp just after t1, and by an incredible coincidence a person, Swampy, is created by this event, who is just like you were at t1. Sequentialism entails that Swampy has what matters in survival for you." [Campbell, 2005. p. 389]

> "Random, being newly created, is not yet causally connected to anything. Replica, on the other hand, is causally connected to things, through being the causal product of earlier states of myself. Thus, Replica can refer to things, while Random cannot (and the same applies to Swampy)" [Campbell, 2005. p. 390]

> "The externalist will then claim that no person can have what matters in survival for me if they do not have any beliefs, quasi-memories, intentions, and so on, because it is these sorts of states that the psychological theorist holds are what matters in survival. So Random cannot have what matters in survival for me. Hence, what matters in survival can exist only if there is a causal connection,

> because without a causal connection there will not exist any beliefs, intentions, etc." [Campbell p. 390]

To counter this objection, Campbell calls the non-causally linked internal states "Near-beliefs" and says that "Near-beliefs will be almost the same as beliefs, and near-intentions will be almost the same as intentions" - Therefore they should be treated the same. I believe that so long as the beliefs are called "near-beliefs", the example will fail, because we talk in terms of beliefs and not near-beliefs. For a start, it is not about whether the beliefs are beliefs or not, but whether they are veridical or not. Whether they are veridical or not depends on whether we accept the non-causally linked person as the person or not. Thus, it is a circular argument.
For instance, if a sequential (but random) being comes into existence, then when they talk about what they did yesterday, we might say that they have no right to speak in such a way, because they did not exist yesterday. But if the only question of whether they existed yesterday is the externalist objection, then that is no proof that they did not exist yesterday.

In Davidson's Swampman example, it is taken as given that the Swampman is not a person, and therefore they cannot have beliefs. However, the Parfit informed viewpoint is that it would be a person, based on understanding what matters. Scott Campbell makes Parfit's view logically consistent with regards to cause not being needed, but runs into problems that only content similarity - particularly wide content via network identity can solve.

In creating Psychological Sequentialism, Scott Campbell has laid a trap for those that oppose him, in that they accept what he says and then try and find an example which is that sequentialism and which seems intuitively absurd. Instead, they would be much better to use my selection criterion in the first section and say that random coherence does not mean anything until it is brought into the meaning sphere via selection by a causally continuous agent. This method could even go so far as to say that the environment creates causal reality when the random agent aligns with it.

We have seen in this chapter that arguments that we need causality for what matters in survival fail and that Scott Campbell has a rigorous example that takes out most arguments for causality being important in survival. Next, we will see if continuity alone is sufficient for what matters in survival.

Chapter 5 - Continuity problems with Relation R

When speaking of psychological continuity, I speak of that which is psychologically continuous, as opposed to something that must have causal connection. In any case, causal connection is knocked out by Scott Campbell's version of the transporter argument. He argues in a 'digital version' of the Swampman argument, that once we are converted to discrete digital information, it can hardly matter about that information's causal history, but rather about it being identical with our current information. Scott Campbell came up with Psychological Sequentialism as the more logical version of psychological continuity, and so I speak of Psychological Sequentialism conceptually when I speak of psychological continuity.

In most ways, I agree with Parfit and Scott Campbell, but I do not agree that what a person is, or what matters to us, is just psychological continuity. On a psychological continuity view, a person has survived if something that has sequential psychological continuity with that person crops up at some point. One problem with this view is that if it were true, we could never know whether a person had survived or not, because we can not know whether or not something with sequential psychological continuity will crop up at some point in the future. So there has to be a further fact about what matters in survival, and that further fact is our connections with others.
Moreover, if a person with a different history should adopt psychological continuity with a prior person, then the original person would not be considered to have suddenly reappeared (the crossover problem).

The crossover problem

The "crossover problem" is my name for what happens if an entity should become psychologically continuous with another entity. For instance, by all of that entities' mental content becoming the same as another entity, so that from then on, the second entity is psychologically continuous with the first entity. For example, if we imagine person "Ted" dies and then immediately after their death we find that someone else "Gene" is mentally the same as Ted was at the time of their death. In that case, it seems we ought to treat Ted as not being dead, but regard "Gene" as their survivor.

The reason that this is a problem, is that if the criterion for what matters is that a person has psychological continuity and connectedness with a prior person (which boils down to content similarity), but if this is achieved by two persons coming into convergence mentally, for whatever reason, it still would not be as good as ordinary survival, because they would have different histories with different rights and responsibilities arising from those different histories.

I consider the crossover problem to be a key problem with psychological continuity as it relates to identity. However, there are plausible and implausible versions of this problem. Noonan gives the implausible one in his book "*Personal Identity*"[46] and it is kind of a 'straw Parfit' - or we could say 'Strawfit'. Noonan says:

> If, say a child reads the autobiography of a historical figure. Napoleon or whoever and subsequently becomes deranged and imagines that he is the person in question, acts and talks like him etc., then so long as no better candidate is around he *is* that person. (Personal Identity, p. 170)

(Parfit would almost certainly disagree.)

> …This is I take it reductio ad absurdum of Parfit's position on personal identity. (Personal Identity. p. 171)

Indeed it is, but more than that, it is a straw version of Parfit's theory. Noonan is saying that mere delusion by itself is enough for personal identity. Delusion in and of itself is a state within an identity state. In this case, it is a child who read about Napoleon, and so it is still that child. Furthermore, it is surely impossible to get enough information from reading about someone to become that person.

The plausible version of the crossover problem is that if a person has the extraordinary delusion that they are Napoleon, and their mental state becomes continuous with Napoleon just before he died, would we then say that Napoleon has survived? Clearly not. Napoleon was a French military leader that was his social and network identity. How can a mentally ill person become the object of their delusion for a few seconds and then stop being that person a few seconds later when their brain states slip out of connectedness with Napoleon's?
There are two kinds of sequential psychological continuity that can follow from a person's death:

1) Psychological continuity and connectedness are possible with a currently dead person because they stopped having brain states at some point. So if someone stops having mental states, they have Psychological continuity and connectedness with all deceased people. This minimalist conception generally is not what people mean by Relation R though.

2) Psychological continuity and connectedness are also possible with a currently dead person if it is the next psychological state which they would have achieved it they *had* lived.

---

[46] Noonan, Harold W.. *Personal Identity*, Taylor & Francis Group, 2004. *ProQuest Ebook Central*, https://ebookcentral-proquest-com.ezproxy.otago.ac.nz/lib/otago/detail.action?docID=182298

In either case, this means that if someone can merely achieve accurate psychological (sequential) continuity with Napoleon at the time of his death, then they become that person regardless of what their future psychological states are, as all future states now have psychological continuity with the achieved state of psychological continuity at one point. So if a person went into a fever and achieved psychological continuity with Napoleon on his deathbed, then, when they recovered from their fever, someone might remark to them, "Napoleon - you speak English without an accent!".

With regards to Napoleon's last words, they would then no longer be his last words:

> Napoleon's Grand Marshal, General Henri Bertrand, did hear some last words early in the morning of May 5th.
>
> From three o'clock until half-past four there were hiccups and stifled groans. Then afterwards he moaned and yawned. He appeared to be in great pain. He uttered several words which could not be distinguished and then said 'Who retreats' or definitely: 'At the head of the Army.' [47]

So someone might ask the person, "Napoleon; what do you mean by: "Who retreats" and "At the head of the Army?" Then whatever they said in explanation would be 'what Napoleon meant by that'. This is clearly absurd.

The absurdity is, that once they have recovered, although there is psychological continuity in a closest continuer sense, there is no psychological connectedness for there to be psychological continuity in a content similarity sense (for instance, memories). There is certainly not enough narrow content similarity to gain network designation as Napoleon.

Skipping a portion of continuity

Another argument against psychological continuity alone, is that the actual continuity does not intuitively matter much in terms of overall identity anyway - much less in terms of what matters. Let us imagine that tomorrow someone wakes up without apparent psychological continuity, by having memories of a day that no one else has experienced. We would definitely say that it is the same person in that case - but that they have some delusional memories. But what happens in a teleportation or swampman example when the person has extra memories? Then they are seen as no longer psychologically continuous, and thus cannot be identified as being the person. We see here (when we compare the two cases) that exact psychological continuity does not matter at all, as long as the person's body is in the appropriate geographical location at the appropriate time. This understanding takes us back to animalism.

---

[47] Selin, Shannon, *What were Napoleon's last words?,* Web, Accessed 22/6/20, http://shannonselin.com/2015/05/napoleons-last-words/

In the Groundhog Day situation that the protagonist Phil finds himself in in the Movie Groundhog Day[48], whatever the memories and personality traits that Phil displays when he wakes up each day, other characters regard him as Phil, which counts against psychological continuity and even connectedness. We might think it does not count against psychological continuity because he has psychological continuity from our (the viewer's) perspective. This is true, but from the perspective of the people interacting with him, he does not have psychological continuity, yet they always regard him as Phil. This case will be discussed in depth in chapter 8.

If we can imagine a further case of teleportation whereby the Enterprise's captain beams away to a planet and then beams back again, he is always regarded as the captain regardless of the underlying process, so long as there is content similarity. Actual psychological continuity is not required. Imagine a baseline case where he is beamed to planet A and then straight away beamed back again. In that case, everyone regards him as the captain. If he is sent back in time, so that he experiences an extra day and is then beamed back, everyone still regards him as the captain. If he is *not* sent back in time so that he experiences an extra day, but somehow he has the mental consequences as if he did, i.e memories and such, the crew would still regard him as the captain. Basically, regardless of what happens, he is regarded as the captain, so long as he is very like the captain in personality and motivations, and we have got an understanding of where his body is in space and time at any particular point.

In terms of our psychological or intuitive understanding of identity, whoever is around that is very like the person *is* the person, and when it comes to problem cases, it is whoever has the same history as per being the closest continuer.

All of these issues tend to suggest that there is more to survival than simply psychological continuity. Originally my answer to this question was "extended self", as in what is the same (content identical) in you as other people, then I came upon the idea of network identity being the predictive connections between ourselves and other people.

Life trajectory

A similar idea to my idea of network identity has also been had by Heidi Savage with her idea of "life trajectory"[49], that being:

---

[48] Ramis, Harold, Trevor Albert, C.O Erickson, Danny Rubin, Bill Murray, Andie MacDowell, Chris Elliott, Stephen Tobolowsky, Brian Doyle-Murray, George Fenton, Pe.mbroke J. Herring, and John Bailey. *Groundhog Day*. Culver City, Calif: Columbia TriStar Home Video, 2002

[49] Savage, Heidi, "What Matters in Survival: The Fission Problem, Life Trajectories, and the Possibility of Virtual Immersion", *Phil Papers*, 2017. [Hereafter "Savage", date and page number]

> "The hypothesis is that the preservation of what matters depends, not just on continuity of the subject's psychology in an internalist sense, but also on the continued living of that subject's life, which involves being embedded in a particular objective environment. It is this continued living of a life that we will understand as constituting a person's life trajectory." [Savage 2017 p12]

> "…to have what matters in survival, a previous psychological subject must be q-continuous with a current psychological subject both with respect to their psychologies and their extrinsic properties." [Savage 2017][50]

Savage soon amends the idea of it being *any* extrinsic properties to being the subject's permanent extrinsic properties. She uses such examples as the Nozick's Experience Machine to argue that if someone was immersed, and virtually recreated in such a machine, although they would have psychological continuity, it would not count as being a valuable continuation of life.

> …many of us, I suspect, would recoil in horror at the possibility of entering the Experience Machine, viewing the persistence of the thing in the machine as a mere simulacrum of what that person once was, and viewing her fate in entering the machine as a fate in many ways like death. [Savage 2017]

So here the question is not; have I survived? But have I survived well or, in Parfitian terms, "as good as ordinary survival"?

Arguments against the experience machine are easy to challenge. If there was a paralysed or locked in person, they might view the option to use the experience machine as not as good as ordinary life, but many would no doubt regard it as a great comfort. Furthermore, there is a character in The Matrix[51] who not only wants to stay in the Matrix, but wants the knowledge of being in it erased. "Ignorance is Bliss …And I want to be rich - someone important…" [Cypher, The Matrix]. This could not even make sense under Savage's viewpoint. Savage counters this by saying about virtual immersion scenarios:

> …consider the anticipation we might feel if all psychological subjects could rid themselves, en masse, of the shackles of bodily decay by immersing their psychologies within a virtual world in which they are maintained independently of their bodies. [Savage, 2017. p. 23]

The idea here, is that if it is everybody together in the machine, that will preserve the social links, when with individuals separately in the experience machine, they are not preserved.

---

[50] Since this paper was removed from the website and replaced with a completely new version of the paper I cannot now reference the page number.

[51] Wachowski, Lilly, Lana Wachowski, Joel Silver, Keanu Reeves, Larry Fishburne, Carrie-Anne Moss, Hugo Weaving, Joe Pantoliano, Bill Pope, Zach Staenberg, and Don Davis. *The Matrix*. Burbank, CA: Warner Home Video, 2001.

The main issue with Savage's theory of extrinsic properties is that extrinsic properties alone do not give us enough purchase to allow for the fulfilment of her original hypothesis of what matters. Content similarity does. It goes back to an issue raised at the start of the thesis: Why content similarity and not extrinsic properties? I believe that content similarity is a more robust and physicalist understanding. There are fewer questions about it. If a person goes out of existence, the creation of another person who has the same mental content as the previous person is all that is needed. With properties, most of them are historically based, and many permanent properties at issue can be stripped away in the five-minute world example, or causality gap examples given previously.

For instance, I have the property of having received a degree from the University of Otago, and I have the property of remembering receiving that degree. A replacement being could not have those properties, but it could have the same mental content.

Consider the example of being the sole author of a particular paper. Savage says:

> …suppose you possess the property of being the author of a certain paper. Then on the life trajectory theory, anyone who can continue your life trajectory must possess that property as well. But only the person who actually wrote the paper can possess that property, and that person is you. It seems, therefore, that any later subject who legitimately continues your life trajectory, which includes the property of having written that paper, must be identical to you. [Savage, 2017. p. 16]

The issue here is that under our ordinary understanding, properties would be wiped out if we were using the Swampman example, or the five-minute world example, and that is just the kind example where someone *could* continue a life trajectory. Moreover, in the Swampman example, whether the Swampman is identical to the previous subject by having its properties is the very question. If we use content similarity as the decider of what is or is not the person we can slot in and out various components within the network, and so long as what we are slotting in and out remains identical, or very close, then what matters is maintained.

Therefore, Heidi Savage's 2017 account, while being one of the best accounts available of what matters, and somewhat close to my view, does not give a sufficiently compelling and detailed account of that subject.

Life trajectory Mark II

In 2020 Heidi Savage published a new version of the life trajectory theory.[52]

---

[52] Savage, Heidi, "What Matters in Survival: Self-determination and The Continuity of Life Trajectories", *Phil Papers*, 2020. https://philpapers.org/rec/TIEWMI

> While I accept that psychological continuity is fundamental for having what is important for survival, I reject the idea that retaining personal identity over time can be exhausted by the holding only of internal relations between a person's psychological states. And, the reason is that such accounts allow for the loss of the property being a person, which I argue is important for survival. To remedy this, I offer a theory I call the "life trajectory" theory that includes satisfying certain externalist constraints, and an address this issue, which most theories of psychological continuity fail to do. [Savage, 2020. p. 2]

In this version, she attempts a much more ambitious project, which does not merely supplement Parfitian psychological continuity and connectedness "r-relation" with life trajectory, but replaces it with the life trajectory theory. It is a two-step process. First, she has to prove that r-relation is not what matters in survival, but life trajectory is. She attempts the first step by using Nozick's experience machine thought experiment. In this case, it is not that we value more than an empty life of mere pleasure which she argues, but that psychological states themselves are not of value, since a person has psychological states *within* the experience machine and yet their life is not a valuable one.

Savage goes further than that and claims that 'people' using the experience machine are no longer persons as they are no longer able to act with self-determination in the world:

> …as beings with self-determined lives, and therefore as being required to be able to interact with and affect their own environments. [Savage, 2020. p. 15]

Why the need to go to the extreme length of claiming that people in the experience machine are not even persons? The reason is that she does not want it to be merely the case that survival consists of Relation R and certain external relations, because then it could be claimed that those relations are merely desirous for different grades of survival, but not essential to survival itself. Realistically, this is to create a hard cut off between all the things that might matter in survival and survival itself. For instance:

> Clearly, there is a distinction between what matters in survival and a life worth living. [Savage, 2020. p. 9]

Savage says that "our psychologies matter, but only insofar as they allow us to continue to be the self-determining beings we had always been." [Savage, 2020. p. 20]. So this gives a very different take on the earlier quote saying, "psychological continuity is fundamental for having what is important for survival" [Savage, 2020. p. 2]. She is saying it is fundamental, but only in an enabling sense. I could also say my legs are important, but only in the sense they allow me to walk.

In this case, it moves the centre of value from the personal to the social.

Finally, we come to the life trajectory theory, which Savage approaches as a non-identity theorist. Because of this, she starts with an argument about intrinsic properties and q-intrinsic properties, which an identical person would have if placed into the same situation. 'q' in this sense is quasi. Savage's argument mirrors Parfit and Shoemaker's argument about quasi-memories.

> A later subject B has the q-property corresponding to an earlier subject A's permanent extrinsic property just in case B is able to relate to their current external context in a way indistinguishable from the way in which A did at the time of their cessation. [Savage, 2020. p. 18]

Savage gives an example of someone having the property of having written a specific paper and says that for there to be what matters in terms of the continued existence of someone having written such a paper, or for there to be a good enough continuer to quality as that continuer being someone as good as that person for the purposes of having written the paper; they would have to have a specific list of internal mental properties, e.g:

> To satisfy condition (1), then, B would have to (a) quasi-believe they were the author of that paper; (b) quasi-remember having authored it to the extent that A did; (c) have the same quasi-attitudes towards the paper that A had; (d) be received in the external world in the same way as A was with respect to that paper, and so on. [Savage, 2020. p. 9]

These criteria are surely mistaken, because either being able to relate externally in an indistinguishable way is important, or the internal mental properties are. Both could be, but if the only purpose of the internal mental properties is success in being able to relate externally, then it is not hard to imagine circumstances in which someone could produce the external outputs, but not have the internal mental content that resembles the previous subject's mental content. They could be a philosophical zombie or a shape-shifting alien that pretends to be various people.

Even more damningly, they could even have the *same* mental property without having the same mental content. For instance, they might have the same property of remembering, but the remembering itself could have a different internal construction.

Two builders working on building a house will each have a memory of having constructed the house - but different memories. In this case, we cannot even say that they need to have the same memory as the person they are continuing from to qualify as having what matters about the person, because that is the very kind of thing that's being disputed!

The important point here is that dropping in someone with a different mentality to oneself in order to continue one's life is a good deal less appealing than dropping in someone with the same mentality. No one wants to have their place taken by a shape-shifting alien who can do everything they currently do. People would rather go into the experience machine than that.

The experience machine example successfully shows us there is more to life than psychological states, but there is a gaping hole in the argument of trying to prove that subjects using the experience machine are no longer people. Below is Nozick's best representation of the experience machine:

> We care about things in addition to how our lives feel to us from the inside. This is shown by the following thought experiment. Imagine a machine that could give you any experience (or sequence of experiences) you might desire.When connected to this experience machine, you can have the experience of writing a great poem or bringing about world peace or loving someone and being loved in return. You can experience the felt pleasures of these things, how they feel "from the inside." You can program your experiences for tomorrow, or this week, or this year, or even for the rest of your life. If your imagination is impoverished, you can use the library of suggestions extracted from biographies and enhanced by novelists and psychologists. You can live your fondest dreams "from the inside." Would you choose to do this for the rest of your life? If not, why not? (Other people also have the same option of using these machines which, let us suppose, are provided by friendly and trustworthy beings from another galaxy, so you need not refuse connecting in order to help others.) The question is not whether to try the machine temporarily, but whether to enter it for the rest of your life. Upon entering, you will not remember having done this; so no pleasures will get ruined by realising they are machine-produced. Uncertainty too might be programmed by using the machine's optional random device (upon which various preselected alternatives can depend).[53]

Even Nozick is not claiming that the experience machine is not giving the *person* experiences. If he was, then it would not be an experience machine. Nor realistically could he claim that that it was not giving the person *experiences*. If the machine did not give the person *experiences*, it would not be an experience machine either - it would be a delusion machine. Compare the difference between a machine that makes someone think they have heard a particular musical piece, to one where they actually hear the music. In the former case, it may be because of a mislabeled disc and in the latter, the correctly labelled disc. If someone plays a mislabeled disc they might think they have heard Beatles, but instead it is the Rutles - That is no good. There is clearly less value in a delusion machine than in an experience machine.

So with the fulfilment of the two conditions of the experience machine: Providing a person with experiences, and with those experiences being provided to a person, how can the subject not be a person? If it is them that is receiving the experiences, then how can it not be a person that is? How can it be claimed that they stop being a person by virtue of receiving certain experiences? If there is any doubt about this case, we can go back to what Papineau said about Swamppeople: "if they have no mentality, as teleosemantics implies, then it would seem to follow, absurdly, that it

---

[53] Nozick Robert, *The Experience Machine*, Excerpt from The Examined Life (1989), Web, Accessed 22/6/2021, https://rintintin.colorado.edu/~vancecd/phil3160/Nozick1.pdf

would be all right to kill Swamppeople and eat them as meat."

Here, we simply need to ask: Can we kill someone with a VR headset on and eat them as meat? As much as we would sometimes like to do that, a quick look at the legal code would suggest that it is still not permitted. In the case of the experience machine, it seems like we would both prefer it to death, and we would seek to preserve the lives of people hooked up to the experience machine. Therefore, I claim that the experience machine example fails to prove that a person's psychological states do not matter. It helps prove that they are not the *only* thing that matters but certainly does not prove that they don't matter.

Overall, what I can say about the Heidi Savage papers on what matters in survival is that it displays great intelligence and creativity, but in seeking to grapple with these massive philosophical issues she has overreached and fallen into a couple of potholes. The example of the experience machine is particularly incredible, since it seems obvious that we would still consider someone hooked up to an experience machine to be a person. As for the failure of the argument about the only requirement of psychological states being to produce external states, it is a little less obvious where that goes wrong, but its refutation is still quite devastating for a theory that is very close to my content similarity theory.

We must then look to see how a network identity theory can solve these problems.

Chapter 6 - Network identity as what matters

For survival to be as good as normal survival, it *requires content similarity over time*. The kind of content that matters in survival is content to the widest extent possible, but specifically narrow content and network content. Here we will explore and detail different kinds of content and their role in what matters in survival.

Narrow content, wide content and network content

Narrow content is the internal content of the thing or person at hand. Wide content is the broad content information in the world overall to a specific scope. For instance, we might duplicate a single human being, while on the other hand, we might duplicate Earth. In the case of duplicating Earth, more content that pertains to a single human being has been duplicated.

There is a reason that I say 'Network identity', rather than simply 'wide content'. A specific kind of broad content is the extrinsic links between a person in question and the other parts of the world, as well as the information held about a person outside of themselves. Network identity also includes beliefs about an artefact or entity that are not truth-apt. For instance, let us say there are two (theoretical) swampmen that have the (correct) added belief that they are swampmen: One tells everyone he is swampman, and a second swampman keeps it secret. In that case, the secretive swampman is overall more like the original man than the 'open' swampman, but that is just because people may have irrational prejudices about swampman as a swamp man. Irrespective of the narrow content, these beliefs may lead more quickly to an erosion of the network content similarity of the swampman if people act on such prejudice.

Network identity as a subset of wide content

I define "network identity" as a limited kind of wide content that applies specifically to an entity and is specifically linked to that entity. It is wider than social identity but narrower than being all content. I see network identity as being a prime candidate for a 'further fact' about identity and what matters. Network identity is other people's and other entities' understanding of the identity of a particular being and their tracking of that being through space and time. The defining attribute of a network identity element, as opposed to a general wide content element, is whether the content involves information held within the network regarding the subject in question.

For instance: It would seem to be enough in narrow psychological content terms to have a ghost or a 'brain in a jar', for it to be the case that a particular person persisted. But this would not be as good as ordinary survival. In the widest content terms, if the entire universe was duplicated, then

this definitely would be enough for ordinary survival, but this leaves no room for any non-standard continuation of survival. Network identity as part of wide content takes what is important to us - what matters to us, and preserves that at the very least. Network identity also provides part of what matters in terms of the distinction between identity and what matters about future states.

A person's network identity is in the thoughts other people and entities have regarding a person, and any information about that person held in any form in the world. One way to understand it would be to understand what the effect of complete and ongoing erasure of network identity would be like.

The effect would be to be a forgotten ghost. For example, the psychological disorder of depersonalisation describes what it is like to have a perceived separation from your self:

> I feel like I don't exist, like I am a ghost who is walking round and interacting with the world my presence is not acknowledge and I have been forgotten by everyone.[54]

This is a statement by someone in a depressive state. I believe that this statement conveys some meaning to us. If there was nothing more to existence than psychological continuity and connectedness, then what information could be conveyed by such a statement?

Having your existence forgotten would be in many ways a fate worse than death, as at least with death some people remember your existence. With the erasure of network identity, there is not even that. In the case of an ongoing network identity elimination, no information about you would be able to reach anyone else. They could not see you or hear you, and nothing you did would have any effect on them. Understanding how devastating this would be to the average person is to understand a massively overlooked part of identity and what matters in survival.

Wide content, further facts and final facts

Let us briefly divert back into what identity is and away from what matters, to see the effect of wide content on what identity is.

Parfit proposes that there can be no further fact about personal identity other than Relation R: psychological continuity and/or connectedness. This view is essentially content similarity about psychological states, which he makes clear when he talks about reductionism. Parfit writes:

> When we describe the psychological continuity that unifies some person's mental life, we must mention this person, and many other people, in describing the content of many thoughts, desires, intentions, and other mental states [R&P 210]

---

[54] scienceguy, *Depression Forums*, 2015, Web. http://www.depressionforums.org/forums/topic/110543-feel-like-i-dont-exist-anymore/

There are two ways in which there could be a further fact about intuitive identity, which is also a further fact about what matters.

1) Social beliefs about identity

Regarding my identity; My understanding is that other people's opinions about it, and our own opinions, can form a further fact about what matters in survival. Let us say, for instance, that no one thought that personal identity was maintained by an unusual cause such as a teleporter, or by no cause at all such as swampman. In that case, it would form a reason to avoid teleporters and swamps during lightning storms. It is an obvious reply to say that people may be wrong in their opinions, but since identity is at least a partially social phenomenon, it does matter what people think about it. It is a fairly obvious point that animals are animalist.[55] People may not be. People's understanding of identity has a great deal of influence on their identity itself. People may have been much more willing to undertake a brain transplant in ancient Greece than presently, due to the ancient Greek belief about the location of the thinking part of the body, which they thought was the heart. Therefore, it is clear that identity, as socially mandated, may require much more than just Relation R.

2) Network identity as a further fact about what matters in survival

Parfit opposes the 'further fact' view about personal identity. That is, that there is a deep or further fact about personal identity beyond physical and psychological continuity, such that an abnormal cause of physical and psychological continuity could fail to produce identity, such as the case of teleportation.

Strictly, there might not be a further fact about 'personal' identity where narrowly defined as pertaining to the person only, however when it comes to the concept of survival being "as good as ordinary survival", the further fact exists in the domain of the social and network identity of the person in question. If the non-normal cause does not preserve that, then the person's overall 'identity' will suffer. So there may be no 'final fact' about identity (due to non-identicality in entities over time), but Parfit is wrong about there being no further fact about personal identity. That further fact is the network that exists as external connections between other people and us, i.e. our 'network identity'. This can be disrupted by abnormal causes, thus forming a kind of further fact. For example: If someone said they had 'died' during surgery, and we took them at their word and treated them as a new person without the rights and responsibilities of the previous one, we would

---

[55] And we regard animals in an animalistic way in terms of personal identity. e.g. a snake with two heads, whereas a human with two heads is two people.

have the kind of conflict that arises with the Swampman case or the teleportation case. In those medical cases where someone says, “I was dead for five minutes”, our intuitions are almost universally with the continuance of the person who ‘dies’. In other words that they do not die, and their identity is preserved.

Where there is a dispute about identity involving a living being, such as in the Swampman case, or even the teleportation case, we can seek to apply social or legal death to the disputed person or not. But a very important component of this is whether their identity is even in dispute or not. The very fact as to whether people believe that identity is extinguished or not provides a ‘further fact’ as to whether it is as good as ordinary survival or not. The capacity for this to be disputed is not even necessarily based on science but may even be on a supernatural basis. In any case, whatever its cause, the dispute itself provides at least a social ‘further fact’. More widely, the external connections between other people and ourselves provide a further fact regardless of the cause of their disruption. There are various ways in which these connections might be broken. It might be just that an extended amount of time has passed, and the persons who hold those external connections have themselves died off, such as in a time travel case, or it could be Nozick’s “a very improbable random event occurs elsewhere in the universe”[56] random coherence case. In the latter case, the fact that it happens elsewhere in the universe causes a severing of network connections, allowing Nozick to generate the intuition that it is not us. If we are instantly replaced, as in the swampman case, then the intuition is far less strong. Furthermore, if we do not think this is a further fact about our overall identity, then our deaths are always contingent upon reanimation at some later date. In which case, we never know whether or not someone has survived.

Examine the iceman case where a mummified body was discovered by German tourists and then pulled from a glacier.[57] If we reconstitute the iceman as he was at the time of death, has he then survived? If he *has* survived, then how can it be that the intermediate existence of his mummified body somehow allows the experiences to belong to the same being, whereas they would belong to a different being if the mummified body was not present?

The mummified corpse is certainly irrelevant to the iceman’s continued existence because there is no final fact about the iceman’s continued existence even when he was apparently alive. There is the closest continuer understanding of intuitive identity, but even then, what appears to be the iceman from one second to another cannot conclusively be said to be the iceman.

The interesting thing about the mummified corpse in the glacier situation, is that under our intuitive understanding of identity, we do not have a point when we can conclusively say that the person’s existence has finished - it is always subject to future events. If, for instance, we could miraculously revive the iceman’s corpse, then the person we would naturally consider him to be

---

[56] Nozick, Robert. *Philosophical Explanations*. Cambridge, Mass: Harvard University Press, 2003. p. 41.

[57] “The mummy”, *Ötzi, the Iceman*. South Tyrol Museum of Archaeology, Web. http://www.iceman.it/en/the-mummy/

would be ‘the iceman’. This breaks our intuitions about identity much more than the split-brain and the teleporter example. If someone gets trapped in a glacier for years and we later revive them, then under our current intuitions, that is not a new person. But where it is someone created ‘from scratch’, straight after atomisation of the original, then many would consider them as such.

We can see that a good part of the further fact about overall identity has to do with whether people believe, or do not believe that someone has, (or does not have) some or other identity. But it has the obvious relativity that just believing it makes it true.

If we think about ‘identity theft’, then we have the situation where people external to us can ‘steal’ our identity. That is strange given that we typically do not believe that identity has anything to do with those external factors, and yet we believe that a change in the understanding of them amongst other people can lead to a situation of ‘not as good as ordinary survival’ if our identity is completely stolen. It is somewhat contradictory to hold these two beliefs about 1) The lack of importance of external factors to identity, and 2) It being crucial at the same time.

I see two kinds of external identity factors that might be at play. One is the internal understanding - who you think you are, and the other, the external understanding - who other people think you are. So in the case of transformation, you might not believe that you survive teleportation, or lobotomy or taking LSD or whatever, and you are a new person now. The external account would be similarly prescriptive as to whether a specific identity was rightly applied and/or at issue in the absence of anything else. For instance, if it was simply the case that society regarded teleported persons as entirely new people, with the previous people eliminated from existence, then that social fact alone would cause teleportation to not be as good as ordinary travel. In other words, a good part of identity may be ‘what we think about identity’ This is already the case with things like skin colour and ethnicity in general, but I would propose that the understanding be widened out to include all understandings of personal identity.

The other part of what I call network identity (that is relevant to our intuitive identity) is how we perceive other people to perceive us. This perception is not simply in identity terms of whether we are such and such a person or not, but rather how we fit into the overall social schema. This is very simple and basic. It can be as simple as going on holiday, which causes a pleasant feeling of identity change.

Alternately, it can be something like feeling a loss of identity from being a refugee, or in otherwise forced circumstances. People might say they have ‘lost their identity’. Yet none of this is seen as relevant when it comes to the philosophy of personal identity. Nothing outside the person, or even the person’s own view is seen to matter. So if there is a swampman that replaces me, then I might be thought to no longer exist, whereas if a corpse is shot into space and revived in a million years, I might be thought to have survived. Surely it must be that our identity in our social and societal network is at least as important as our bodily identity.

This brings into focus an issue about identity, that a good part of our normally understood identity is external to us. Whether this is 'personal' identity or some other kind of identity is not widely considered. Someone can 'steal' your identity and 'invade' your personal space. If we consider our wider social or societal identity to be part of our personal identity, this creates many problems for traditional theories of personal identity. We believe there can be "identity theft", but can someone steal a person's personal identity? One response to this could be that if someone were really to steal your personal identity, it cannot be them that is stealing it. When we speak of someone 'assuming someone else's identity', they are not putting on their body like a suit, nor have they become psychologically identical to the person. It means that the assumer has taken the other person's place in the public network of connections normally occupied by the person that they have assumed the identity of. Consider the following case: There is a murder. The murderer feels very remorseful about his actions and so what he does is construct a machine that will turn the molecules in his body into a duplicate of the person he had killed, just before he killed them. Should the friends and colleges of the victim accept the duplicate as being the murderer or the victim? The resulting being may have a particular view of their personal identity, but this will be less important to their life and ongoing happiness than what other people think.

I have shown in this section that we can see network identity as part of personal identity and certainly part of what matters in survival. In the next chapter, we can see direct proof that persons require extrinsic factors beyond Relation R, for not just as good as ordinary survival, but simple survival.

Chapter 7 - Interactivity in network content

In this crucial section, we see what is wrong with Parfit's argument for what matters in survival. I do so at first in a fun way and then in a more serious way.

The scholar's mate on Parfit

Parfit says that "what fundamentally matters is Relation R (psychological continuity and connectedness) with any cause" [R&P 217]. In the popular conception of a ghost, it has Relation R in that we can imagine it having a mental life, yet by definition, they have not survived. They certainly do not have as good as ordinary survival. What about a blind, deaf ghost? That sort of entity could have Relation R, and that would be a fate worse than death.

That is actually 'mate in one move'. Parfit has said what matters in survival, and straight away, I have given an example that meets those conditions, and yet the entity concerned does not have what matters in survival. This is because it is actually a supposed outcome of someone who in the popular imagination has not survived. Ghosts have the identity of the person but not what matters in survival. They are not survivors. Parfit must have heard of the concept of ghosts, so why does he persist with his idea? There are two responses Parfit could give:

1) He could say that a far-fetched supernatural example does not matter. So that if there were ghosts, then that would be a problem, but there are no ghosts, so it is not. After all, Parfit explicitly rejects the spiritual view of souls in the form of Cartesian Egos [R&P 228].

If Parfit gave such a response, then the example could be naturalised into a technological form, such as uploading a consciousness into an isolated computer. Parfit also uses a host of non-actual examples like split brains, fission, fusion, and teleporters.

2) More likely, he would say: "What I really meant is you have got to have psychological Relation R and a body." Parfit defines a person as a mental being in conjunction with their body "Our existence just involves the existence of our brains and bodies…" [R&P 216].

Thus a person has to have a body. In the split-brain example, someone can have half of their brain removed, and they are still the same person. What about if we use the unused space in the head of someone who has had such a hemispherectomy and placed another half brain in there? However, when the surgeon goes to connect up the brain hemisphere to the spinal cord, he cannot be bothered, knowing that that is not a requirement for survival. Or maybe he cannot. Maybe the

technology only exists to connect the blood vessels to keep the brain alive, but not hook it up to the nervous system. - Like now, for instance. So the (half) brain is in a body that is walking around, and it has psychological continuity and connectedness, but we would hardly consider the person to be surviving well.

If we placed a whole brain in the body of a comatose person, but did not 'hook it up', that would hardly be considered any better. This might be something that could currently be done, so why isn't it done? That is, why doesn't the medical community offer the transplantation of a healthy brain into the body of a brain-dead person? It seems like this should be a transplant option for terminally ill people who have a condition that does not affect their brain, and yet it does not seem to be used as a treatment option. They would have a body. They would have psychological continuity and connectedness, so what is the issue?

I propose that the issue is that it would not be a life worth living and therefore not as good as ordinary survival. That was move 2.

At this point, Parfit might sigh again and say: "What I really meant when I said over and over again that what matters is Relation R, is that what matters is Relation R and having a body and having that body be hooked up in the right kind of way so that the person can conduct their affairs in the normal way.

Prior to saying what matters in survival, Parfit says we are not separately existing beings apart from our brains and bodies:

> Our existence just involves the existence of our brains and bodies and the doing of our deeds and the thinking of our thoughts and the occurrence of certain other physical and mental events [R&P 216]

So Parfit first defines what it is for us to exist, then he says that what matters in survival is something other than that - Only Relation R. But it turns out that what *really* matters in survival is that we have our criterion for existence in identity terms be fulfilled at later times. This is a big problem, because the entire point of Parfit's project is to pull identity and what matters apart. That was move 3.

If I were to ask Parfit WHY someone needs to be able to do things, rather than just think thoughts, he would have to say it is to connect with the outside world and have certain kinds of relations with it. So that is what is also important. Those relations are part of survival being as good as normal survival. That is what I mean by network identity.

"There is a reason that I say "Network identity", rather than simply "wide content". A specific kind of broad content is the extrinsic links between a person in question and the other parts of the world, as well as the information held about a person outside of themselves."[58]

That was move four: Checkmate on Parfit - The 'Scholar's Mate' if you will.

Persons remaining an active part of the network

The heart of the problem with Parfit's understanding of what matters in survival is not strictly the breadth of content preserved and the links within that content. More importantly, is whether the subject in question remains an *active* part of the network[59]. When Parfit uses teleporter examples to say that what matters is psychological continuity and connectedness, he silently imports an extra concept: A human body and its functionality.

There are two ways in which a person can be cut off from other people, and yet we could still consider them to have in some sense survived.

Firstly a person could be transported alone to a distant place, and we could still consider that to be survival in a normal sense although not as good as ordinary survival.

Secondly, there could be a scenario where a person survives but no longer has any links with anyone else because of the erasure of everyone's memory of them, so for instance, their friends and family no longer know who they are. So it not being as good as ordinary survival indicates that there is value in the extrinsic world. However, some may maintain that they have still survived in the normal sense and can build up new connections again.

Finally, we come to the case where the person's consciousness survives, but they have no access to the world. At that point, we can no longer say that they are surviving, particularly where they no longer have a body at all. An example of this would be a brain in a jar where the 'jar' is not connected to any means of communication and could not be. An illustration of this would be the situation where a person died, yet their consciousness still survives in some sort of ghostly form, unable to interact with anyone, especially if they are unable to send and receive information between that consciousness and the world at all. In that case, it seems likely that the only response we could have to the death of someone's physical body would be the response we typically have now - that is, to bury or burn their body and to have cultural ceremonies of farewell. In that case, they would have Relation R, but people would presumably act as they do now in terms of bereavement. It could even be worse, because knowing that the person's consciousness was still active could make the deceased's loved ones even more upset. The heart of the problem

---

[58] Chapter 6 of this thesis.

[59] Parfit himself has in some sense remained an active part of the network after his death by me engaging with him in this thesis.

with Parfit's understanding of what matters in survival is not strictly the breadth of content preserved and the links within that content.

This section has proved that something more than Relation R (psychological continuity and connectedness) matters to survival. Next, I seek to explain how Parfit did not see this.

Parfit believed that Relation R was all that mattered in survival, yet he would certainly have come into contact with the popular conception of ghosts and souls, where it is evident that Relation R is not all that matters to survival. So how could Parfit not have taken account of this information? I believe the answer to this question is that Parfit had already sought to distance himself from non-reductionist conceptions of personal identity:

> Many Non-Reductionists believe that we are separately existing entities. On this view, personal identity over time does not just consist in physical and/or psychological continuity. It involves a further fact. A person is a separately existing entity, distinct from his brain and body, and his experiences. On the best-known version of this view, a person is a purely mental entity: a Cartesian Pure Ego, or spiritual substance. But we might believe that a person is a separately existing physical entity, of a kind that is not yet recognised in the theories of contemporary physics. [R&P 210]

Although Parfit takes the Cartesian view very seriously for the purpose of refuting it, he makes two mistakes:

1. He ignores the fact that even where people attest that a person is in some sense a separately existing entity, spiritual and religious people believe that death has more or less the same consequences as non-religious people. For instance, we redistribute the possessions of dead people regardless of our religious beliefs. If we were Christians who believed in the second coming and resurrection of believers, we might otherwise leave the possessions of the faithful to be used upon their return, much like ancient Egyptian burial traditions.

   Essentially, regardless of religious beliefs, we regard people who no longer have a body as no longer existing for legal and most social purposes.

2. The more serious mistake is that after dismissing the Cartesian view, Parfit fails to consider existence aside from our brains and bodies. Partit's view is essentially that what we are, are bodies and brains, but what matters is psychological continuity and connectedness within some body and brain. For Parfit, the body and brain are the essential containers of consciousness, whereas consciousness is the really important part in that container.

With the second of these mistakes, Parfit ignores the possibility that something other than a body and a brain could contain a consciousness. We could theoretically upload a consciousness to a computer, for instance. We can easily apply Parfit's arguments about eyes and artificial eyes to brains and artificial brains [R&P, p. 208].

We must define a person differently from Parfit, that being that a person is Relation R and a means to interact with the world outside of their consciousness. That means of communication is typically a body, but need not be.

To summarise: Parfit defines *survival* as requiring Relation R only. In this section, I have shown that we require Relation R and the kind of connectivity that our bodies give us for a state to be survival. In previous chapters, I have argued that with regards to different grades of survival, specifically survival that is as good as ordinary survival, positive extrinsic relationships and statuses need to able to be maintained.

Part 3 - Time loop examples

Now that we have set up the theoretical structure to understand what identity is, and what matters in survival, we can look at an example where we can draw on this knowledge to make sense of a situation. We cannot draw on this knowledge in normal factual situations because such a situation whereby we might need the knowledge never arises. However, we can use certain counterfactual examples to assess and make sense of a situation that a protagonist might find themselves in. One such counterfactual situation is a time loop.

A time loop is a situation where a period of time 'loops', except that for certain participants their memories of previous loops may be maintained. Our intuitions about participants in time loops can tell us a lot about how we perceive identity and what matters in survival, and what theoretical basis for understanding it makes sense.

Chapter 8 - The Groundhog Day scenario - What it is and why it matters

Imagine the following scenario: One night, Phil goes to bed as usual. When he wakes up in the morning, everything seems normal at first, but as he goes about the day, he soon notices that everything is exactly as it was the previous day. People seem to be doing the same things as yesterday, and they do not remember the day he experienced yesterday. Also, the flow of events is the same, unless he explicitly causes it to change.

This situation repeats day after day with no end in sight. This single repeating day is now his new reality, but should this fact influence his behaviour throughout the day? If so, or not, then why?

This question is at the heart of a number of fictional treatments of the situation, most notably "Groundhog Day", a 1993 Movie featuring Bill Murray.[60] From this, I have shortened the name of the situation to GHDS or "Ground Hog Day Scenario".

It is worth recounting the basic plot of Groundhog Day, since it is the reference point for the many variations that came after, and because it is the name given to the phenomena of repeating the same thing repeatedly.

Groundhog Day (the movie) centres around one day that Phil Conners, a TV station weatherman, has on a visit to the small town of Punxsutawney. He is there to witness the Groundhog Day festival where a Groundhog (also called Phil) is said to predict whether spring will come, or whether there will be six more weeks of winter. When Phil gets to the town, he checks

---

[60] Ramis, Harold, Trevor Albert, C.O Erickson, Danny Rubin, Bill Murray, Andie MacDowell, Chris Elliott, Stephen Tobolowsky, Brian Doyle-Murray, George Fenton, Pe.mbroke J. Herring, and John Bailey. *Groundhog Day*. Culver City, Calif: Columbia TriStar Home Video, 2002

into a guesthouse and goes to bed. He awakes the following day and delivers his TV report on the Groundhog Festival. Phil and the crew attempt to leave but cannot get home because of an unexpected snowstorm along the route out. Phil returns to the guest house for another night, but when he wakes up the following morning, he finds it is the same day again, with the only difference being that he has memories of the previous day; February 2nd, Groundhog Day. This cycle repeats day after day, with only Phil's (deteriorating) mental state being continuous. Initially, he is confused, and then depressed, but then Phil's behaviour becomes increasingly erratic and unhinged until he says, "I'm not going to live by their rules anymore." [Groundhog Day, 33:44] He then turns a corner onto a new hedonistic lifestyle, whereby he indulges himself in many ways.

After some time (in his mental understanding), hedonism starts to wear thin, and Phil sets his sights on forming a romantic relationship with his producer. At first, he uses the 'trick' of having superior biographical knowledge about her, which he has access to from his experience of the repetition, but this only gets him so far, and he is ultimately repeatedly rejected. Romantic rejection sends him into despair, and he repeatedly tries to commit suicide, with the only result being that he wakes up back in his bed at the guesthouse. Faced with this circumstance, Phil turns to developing his personality into one that is less narcissistic and more regarding of others. His more generous persona has the ultimate result that he is able to woo his producer, Rita, and convince her that he is an acceptable life partner, at which point the loop finishes, and Phil exits back into normal narrative time.

Groundhog Day is only one of many different stories based on repetition, and although many of these are simply based on Groundhog Day, it is worth taking the time to at least compare their details with the original and other real-life situations where repetition is a feature.

The chart below provides a brief comparison:[61]

Type of time loop	Brief description
1) Friedrich Nietzsche The concept of "eternal recurrence"	The entire universe is in a time loop. Only some people may recognise this, and realisation or awareness of this possibility is supposed to have major psychological consequences.
2) Groundhog Day	A single man is in a one-day time loop in a small town, which everyone else is unaware of. After a large number of loops he exits the loop.
3) Edge of Tomorrow	A single man is in a time loop during a war. He finds a woman who has also been in a previous loop state, but only he has memories of his previous loops.
4) Star Trek TNG, Cause & Effect	The Enterprise is caught in a time loop due to a temporal phenomena in space. This eventually becomes apparent due to déjà vu being actual memories of past events.
4) X-Files, Monday	The character Mulder is caught in a time loop of which only he has partial memories, primarily a feeling of déjà vu.
5) Stargate SG-1, Window of Opportunity	Earth and another world, that it is connected to via the Stargate, are caught in a time loop due to the activation of alien machinery that needs to be decoded in order to be stopped. Only two members of the Stargate crew have memories of the previous loops.
6) Medical condition: Transient Global amnesia	The subject is caught in a 'time loop where they forget biographical information after about two minutes, and ask the same questions repeatedly - often in an entirely predictable way.
7) Medical condition: 'Groundhog Day student'	A person has a constant and overwhelming sense of déjà vu. The person has insight, but does not do certain activities such as reading newspapers, because he feels he has done them before.
8) Medial condition: "Chronophrenia"	A case where a woman believed she had already lived her life before and had memories of the previous life which influenced her actions in the current life.
9) Doctor Who, Heaven Sent	The doctor is trapped in a transporter time-loop, in a castle designed to get information out of him.

---

[61] 1) Nietzsche, Friedrich W, Thomas Common, Paul V. Cohn, and Maude D. Petre. *The Gay Science*. , 2020.
3) McQuarrie, Christopher, Jez Butterworth, John-Henry Butterworth, Doug Liman, Hiroshi Sakurazaka, Tom Cruise, Emily Blunt, and Bill Paxton. *Edge of Tomorrow*. , 2019.
4) Roddenberry, Gene, Patrick Stewart, Jonathan Frakes, Brent Spiner, Gates McFadden, Michael Dorn, LeVar Burton, Marina Sirtis, Whoopi Goldberg, and Denise Crosby. *Star Trek, the Next Generation: Season 5. Disc 5*. United States: Paramount Home Entertainment, 2002.
5) Anderson, Richard D, Michael Shanks, Amanda Tapping, Christopher Judge, and Don S. Davis. *Stargate Sg-1: Season 4 [disc 2]*. Santa Monica, CA: MGM Home Entertainment, 2003.
6) Larner, A J. Transient Global Amnesia: From Patient Encounter to Clinical Neuroscience. Place of publication not identified: Springer International PU, 2018. Print.
7) Knapton, Sarah. 'Groundhog Day' student trapped in bizarre déjà vu time loop for 8 years. The Telegraph, 20 January 2015. https://www.telegraph.co.uk/news/science/science-news/11356853/Student-trapped-in-bizarre-deja-vu-time-loop-for-8-years.html
8) Pethő, B. "Chronophrenia - a New Syndrome in Functional Psychoses." *Psychopathology*. 18 (1985): 174-180. Print.
9) Nettheim, Daniel, Justin Molotnikov, Peter Capaldi, and Jenna Coleman. *Doctor Who: Series 9. Part 2*. London: BBC, 2016.

One question is why we should even care about any of this? It being merely a fictional scenario among many fictional scenarios. Actually, there is a reasonable amount of religious and philosophical engagement with Groundhog Day's ethical themes. For example, Catholic philosopher Michael Foley says in Touchstone:

> "The ground of Groundhog Day, in other words, is Catholic. And just as our secular celebration of the day unwittingly echoes a deeper truth about the Light revealed to the gentiles, so too does the movie unwittingly point the way back to that truth."[62]

That attention certainly helps the study of repetition and the GHDS (Ground Hog Day Scenario). However, that is not why we should care about it. The real reason is: *That once certain variables are altered, our behaviour seems likely to alter, but those variables were never seen as being influential in the behaviour to begin with.*

The most important thing to come out of the GHDS, is that thinking about how we would act in it forces us to move away from causal history and towards content similarity as an explanation for regarding something as having value.

When people speak of reasons for our beliefs or actions, they typically refer to events that have taken place in the past. I propose that the events *actually* having taken place in the past are not a necessary condition for referring to them, or reacting to them, but rather that the current content of the universe is as it would be if they had occurred. By content, I mean both in terms of substance and its organisation.

My understanding is that the causal history was never necessary, but was merely a convenient place to supposedly draw value from because it is incredibly analogous to the true source of value which is content similarity. In the philosophy of science, there are many examples of such parallels. Florence Nightingale did not understand the modern theory of how and why diseases spread, but she had a somewhat parallel understanding which accorded her success. Moreover, it is almost always only because of causal history that we can have content organised in the right way. It is just that if we remove causal history, the value remains from the content being organised in that fashion, regardless of how it came to be so. Confusing content similarity and causal history is a bit like confusing something with the means of doing that thing. For instance, someone might be Christian and think a bible has value because it contains God's word. Indeed, before the invention of electronic texts, bibles being books was essential, and therefore to get rid of all books

---

[62] Foley, Michael. "Phil's Shadow". *Touchstone*, 2004, Web. https://www.touchstonemag.com/archives/article.php?id=17-03-012-v

would be to effectively extinguish the word of God (for those that believe it is such). These days it would not do so.

A specific state is maintained by informational similarity, not causal history; however, to solve certain problems, it is required that knowledge of this fact is included in the similarity. So a universe created five minutes ago is more similar to a causally antecedent universe where people do not know it was created five minutes ago, than where they do. This more or less solves the causal theory of reference problems that may arise. If the causal theory of reference is true, then different things are true in a state of identical content if they have a different causal history, but this is only an issue where it is recognised to be as such. However, for someone to recognise it as an issue is a difference in the content itself, and therefore it is not an issue.

In the GHDS various ontological explanations could be behind the events. For instance:

1) It could be that Phil is travelling back 'mentally' in time to the previous day.

2) It could be that Phil dies at the end of each day, and then he has generated anew, where he was located the previous day, along with everyone else, but only his memories are updated.

3) It could be that Phil is being transported mentally to a parallel dimension where everything is the same as the previous day in this dimension.

4) Each 'day', we could just be being shown a parallel dimension where everything is the same except Phil's memories are updated.

5) It could be that at the end of each 'day', some machine or force is altering everyone and Phil's brain and body (but not his memories) to be the same as it was at the start of the previous day.

I will not examine them individually, but depending on which one of these scenarios is true, a traditional theory of ethical and general action is thought to be required under a causal theory but not under a content similarity theory.

Let us examine possibility number five. For instance, say Bob enters into a contract so he owes Bert five dollars. Bob uses a machine that alters his brain and body and the universe in general so that the world is the same as before Bob made the promise. In that case, Bob no longer owes the money. That is the intuition that arises from Groundhog Day. For instance, if Phil plays poker with someone in Groundhog Day and then wins, but he is not paid before the end of the day, then that debtor does not owe him any money the following day. Of course, it would be possible to claim that Phil does owe the money but does not know it. I say that it is not reasonable to make that

claim. This differs from forgetting because forgetting is a state where someone had a memory but then lost it. However, in returning a person to a prior state, they effectively never had the memory to begin with. Any cascade of subsequent mental events that flowed from the recording and forgetting of the memory are effectively removed. The key issue is whether the information still exists in the universe. Information can only be held as a state of the universe. If the universe goes back to how it was and there is no further information outside the universe, then someone having a memory pertaining to a previous state of the universe - now inaccessible to us, does not create the rights and obligations that such a memory would normally entail. For instance, imagine that we are in a universe-wide time loop of a single day and have been for two 'cycles', but no one remembers the loop. That should mean under our normal conception of things that a whole lot of extra rights and responsibilities have accumulated during that extra day. But if there is no information regarding that which exists in the world, then those rights and responsibilities can not exist in the time loop world.

Here we have seen that scenarios whereby there is a time loop occurring are important to examine, because they allow us to see that the actual occurrence of something is of no relevance, compared to the information relating to that thing still existing.

Chapter 9 - Identity and obligation in timeloops

Concerning identity in a GHDS, our current intuitive conception is this: Whatever it is that wakes up in your bed tomorrow (and is your closest continuer through the night) that looks very like you, is you. Even if you have had a complete personality and mental change overnight, that person (who may not even have any of the same mental content as you) now inherits your rights and responsibilities. That is our basic animalism under the law. Simultaneously, another intuitive conception is: An actual duplicate of someone, without the same causal history, would inherit none of their rights and responsibilities. For instance, if a swampman replica arises, then that swampman does not gain the rights and responsibilities of the person they are identical to. The intuitions raised by the GHDS are that rights and responsibilities do not work in this way, even if everyone in the situation thinks that they do. For instance: There is a fact about the world that makes it that even if we believe that we are entering into a contract on a certain basis and that basis remains unchanged, then that contract is still not valid. It is not that we cannot enforce the contract - It is not a valid contract.

When Phil wakes up, everyone regards him as they did yesterday despite his changed mental content. They expect the same things of him as they did their previous day, so reciprocally he has the right to expect the same things of them as he did yesterday. But neither of them should. Neither of them has that right. If they *did,* it would simply be his task to convince them of the fact and vice-versa. This is not their respective tasks, and it is not just a question of reasonableness but a question of logic. It cannot be that there are conflicting rights and responsibilities in this way.

If we use the causal history theory to resolve these cases, then “What *actually* happened in the past” is the question that we would ask. This is all very well, But what if the past’s conflict? Who owes what then? If I acquired debt to you of five dollars in my past and in your past you acquired debt to me of ten dollars, but no one remembers any of it, then is any obligation still in force? Who owes who what? The resolution to this problem tends towards a somewhat Lockeian ‘memory’ view of personal identity, but rather than memory or lack of it being the factor in itself, memory is merely part of the overall content being considered. This is if we can even posit the idea of a ‘past’ in that sense. The past can only be the past if there is information pertaining to it in the universe. Where such information is absent, or only represented by some local and unconnected content such as one, or even a number of people’s memorys, then it has little relevance.

What about the role of content similarity in the GHDS?

Content similarity is an issue in the GHDS, in that we would not want to think that ‘particular history’ was such an issue when Phil wakes up each day. That is, how Phil reacts to the world should not depend on the exact metaphysical circumstances of his plight - which is what it would have to be for the ‘particular history’ enthusiast. For instance, if the entire universe from the

previous day had just been duplicated, and Phil had been thrown into it, I would not expect that would be the prime issue as to whether he could continue to pursue relationships with the people in that world.

Surely the issue is the interruption (for him) of normal narrative time, and non-memories of all their connected experiences, that people around Phil have that is the problem - It is not contingent on certain metaphysical facts that he does not have access to, that are the determiners of how he should act.

My claim is that, although the metaphysical or cosmological setup is thought to be very important, it is content similarity within any particular setup that is important.

What objections could there be to such a content similarity thesis?

It might seem that only being held accountable via state of content similarity could be a convenient way to get away with crimes. For instance, someone might decide to steal a certain amount of money. They might then place the money in their house, where they might easily come across it and then step in a machine that returns them to the mental and bodily state of before they had the thought of stealing the money. The person's experience then is of unexpectedly finding money in their house. The question is: Should they be held accountable for this crime? Under a causal history view, the answer is definitely "yes". However, under (at least) narrow content similarity, it looks like the answer is "no" because their content is the same as before they committed the crime, and so it would be equivalent to convicting someone of a crime that they could be accurately predicted to do, but had not done yet.

If we accept the intuition generated by the GHDS, the crime should remain unpunished because rights and responsibilities can not be transferred in that way. Of course, they would not have any right to the money either because they did not have any right to the money in that previous situation. It is just a question of whether someone should be punished for something they did not even know they had done.

If such a case seems problematic for the content similarity view, don't worry because there is an analogous difficulty for the causal history view. Say someone makes an exact duplicate of another person. What rights does the duplicate have? Of course, the standard response is to say that the duplicate does not have the right to the original's causal history. If it did, then it would have a right to all of their possessions. This is even doubly more so if a 'swampman' replica emerges. It is not allowed to take your stuff!

But this is a problem when it comes to the restriction of rights. If someone is in prison and their accomplice makes a duplicate of them outside of the prison to carry on their nefarious purposes, then if we do not grant that duplicate access to the causal history of the prisoner, we have no reason to imprison the duplicate.

Under the causal history view, if a 'swampman' Charles Manson emerges, we have no reason to imprison it if we cannot link it to the causal history of an imprisoned Charles Manson, even if they are identical in every respect.

Under a content similarity view, a swamp Charles Manson could be imprisoned on a narrow content similarity basis because he would represent the same danger to society as the original Charles Manson. He would not be Charles Manson in an identity sense as conveyed by closest continuer, but he would have what matters about Charles Manson point to point via content similarity.

What about contracts and agreements with people in the past under the GHDS?
Let us put an argument in a more formal structure and see what we can make of it. To set up:

1) The things and people I have already interacted with are in the past.

2) If someone rationally makes an agreement with me in the past they are bound by that agreement.

3) This is true even if they do not remember making this agreement.

If all of the above is true, then the following scenario in Groundhog Day must be true:

1) Imagine that Phil buys Rita a coffee, and Rita says, "I will pay you back next time we meet".

2) The next day Phil sees Rita, he asks for the money back.

3) We do not believe that Rita has to pay back the money.

The third part of this triad is based on us believing that Rita does not have to pay the money. Some people might believe that she does. I expect most people would have another intuition, as do I.

A breakdown of the argument: One or more of the following must not be true

1) The things and people I have already interacted with are in the past.

2) If (a) someone; (b) rationally; (c) makes an agreement; (d) with me; (e) in the past

Then they are bound by that agreement.

3) This is true even if they do not remember making this agreement.

The most likely candidate for dispute is either 1 or 2e. Both have to do with 'the past'. It is certainly Phil's past. Is it Rita's past? A causal history view would say that it is only Rita's past if there is a direct causal connection between those events and the present (There is an indirect causal connection through Phil). Whether there is a causal connection or not depends on the metaphysical setup of the situation. It seems clear that Rita does not owe the money. So it just does not matter about the causal connection; what matters is the content similarity.

A content similarity view would say something like: Rita only owes the money if there is sufficient content that would follow on from her state of owing the money. At the very least, her state has to be different than before she owed the money. One formulation could be that she does not owe the money if her state is identical to any of her states before she made the agreement.

What about the situation where Rita goes into the GHDS owing money to Phil? It might be easier to think that she does not owe anything after the first loop. If that were the case, it would likely be the same for all other rights and duties. A content similarity view also allows a way for her to legitimately continue owing the money over and over again, even though in Phil's past, she had already repaid the money any number of times. We would then likely say that she owed the money whenever her content returned to the state it was in before she repaid the money.

It is worth bearing in mind that this is just a 'narrow' content view. There is a lot of content in the GHDS, and some of the content - such as Phil's, is not in alignment with this. In a repeating situation, his mental state is in a state post-which she has repaid the money. His financial state is not.

In Groundhog Day, Phil checks into a guesthouse called "The Cherry Street Inn" for one night, before reporting on the Groundhog festivities. He ends up saying another night when a blizzard unexpectedly comes in, which means that he is in the same place for the beginning and end of the time loop which follows.

How many breakfasts is Phil entitled to during his stay at the Cherry Street Inn? Let us say that accommodation and breakfasts are treated separately at the Cherry Street Inn, and Phil is entitled to two breakfasts during his stay. On the third day, Phil comes down to breakfast. He has already eaten breakfast on the two days previously. To how many breakfasts is he entitled? My answer is one further breakfast at that point. So following breakfast on the third day, he has eaten a third breakfast - he is only entitled to two, so how can that be? There are a number of possibilities:

1) He was not entitled to that third breakfast. (But ate it regardless)

2) It was not his third breakfast.

3) It was his third breakfast, but somehow he is entitled to it.

4) It was only the second breakfast that the guesthouse recorded as serving him.

It is a composite case. It was Phil's third breakfast, but only the second breakfast that the guesthouse has noted serving him; therefore, he is entitled to it.

What is it that makes it the case that someone is entitled to something? First of all, there has to be a relation of entitlement that is created at some point, so if the world ever returns to the state that it was when that entitlement was created, then that entitlement is in force again. This is wide content similarity. So Phil has found the world in the same state as when the entitlement was in force. Therefore the entitlement is in force again.

Someone might say that in terms of breakfast as a composite case (based on content similarity), the world is more similar when the guesthouse lays out breakfast than the guesthouse does not lay out breakfast, yet whether they prepare the dining room in that way does not affect entitlement. I would say that in a minor way, it *does* affect entitlement. The closer the world is to when an entitlement was in force, the greater the force of the entitlement.

What about the situation where a person is entitled to breakfast at the Cherry Inn, and at that time, the Cherry Street Inn is painted blue? Then later, when the person partakes of the breakfast the Cherry Inn it is painted red?
Later again, the Cherry Street Inn has the red paint scraped off and is again painted blue. Does that fact increase the entitlement to breakfast?

Even under a content similarity schema, the colour of the Cherry Inn is unrelated to the breakfast entitlement, as it was never relevant content in terms of a breakfast entitlement, even in normal narrative time. It can hardly be relevant in a time loop if it is not relevant in normal narrative time.

The last example I will give is when Phil checks out of the Cherry Street Inn, having exited back into normal narrative time. How many nights stay is he liable for? Let us say that upon checking out, Phil remarks to Mrs Lancaster (the proprietor of the inn) that although he appears to have stayed two nights, he has been stuck in a time-loop and stayed a thousand nights. If she accepts this, how many nights should she charge him for? I believe the correct answer is two nights, but how can this be if all parties agree that he stayed one thousand nights? The reason is that the content of the situation is such that it is identical to him having stayed two nights, except for his memory of having stayed one thousand nights.

To make it clearer, if we could time-loop the whole world, without *anyone* having memories of it afterwards, and if everyone agreed that it had happened, payments, after the loop had occurred, should not reflect the loop having taken place. So you should not have to pay an extra day's rent or get an extra day's pay, for instance. To do so would be to make a fundamental ontological mistake. The mistake, in general, is that since causality and content have generally tracked together, people have seen them in the same light. Where we take causality and content apart, we can see that causality is redundant as means of determining our behaviour - only content can do that.

If I may give a further example: We can imagine that someone presses a button, and what this causes to happen is that the entire Earth, its content and orbital proximity are reinstated to the state they were in 24 hours ago. The rest of the universe is unaffected. We could then say that the relevant content had changed such that the situation of 24 hours ago is the actual situation. So, who is owed what is in the state that it was 24 hours ago, not where it was just before the person pressed the button. In this case, time and the vast majority of the universe have moved on - it is later in time. Despite this, it makes no sense if you stay at a hotel and get a bill for an extra night's accommodation. The relevant content is appropriate to only one night's accommodation.

For the purpose of linguistic rule-following, it is worth pointing out that where content is identical between an original and a copy, our linguistic usage has automatically accommodated content similarity as the criterion of value. We see this with computer files, whereby if I am working on my thesis and my supervisor asks to see the thesis, he might say: "Send me the file". He does not intend me to pack up the computer and send him that, as that is where the file is located. So the linguistic usage has moved to refer to content similarity where people see it to be identical.

Items and item identity

A good example of how network identity operates is in the following case: A burglar breaks into a house while the owner is away. He takes all of their possessions away, and then in order not to arouse suspicion, he makes exact copies of them and takes the copies back to the house and places them where they were. The owner comes home and does not notice the change. Then the burglar's house accidentally burns down, destroying all of the householders' possessions.

Six weeks later, the householder comes home and finds their house stripped of their believed possessions. They ring the police, who say the burger has come in and confessed what he did. He has burgled the house and also put some other items in their place. Unfortunately, they were unable to recover the original possessions, but they got the burglar to remove the items he left in the house. They hope the householder will now consider the matter closed.

What we see here is that item, and item identity are different. When the burglar put the identical items in the house, they took on the network identity of the original items, even though they were not the original items. As such, the victim of the burglary has effectively been burgled when the police have said that they have restored the situation. Once the householder discovers that the items do not have the right identity to be their belongings, it does not matter because the duplicates have what functionally matters about their possessions.

We can see with item identity that items gain their identity from their spatial arrangement over time. The “Item” being any particular defined object. If we can imagine a number of time slices, if we insert an identical or near-identical object or entity into one of those time slices (and take out the original), it can take on the object’s identity or identity through time. If found not to have that identity, it still has what matters.

We can see this when Phil wakes up in bed every morning of Groundhog Day. He is not identical to the Phil who went to bed the previous (non-looping) night, but he is close enough to take on Phil’s identity. As the Swampman case shows, this is a situation that we may be inclined to grant that someone has what matters in survival if they are identical but without the right kind of causal history. Yet in the GHDS case, Phil is not even identical to his previous self so far as the other people in the time-loop are concerned.[63]

To make this clearer, when Phil wakes up, he is close to identical (mentally) to how he was the moment before in the time loop (along the lines of normal psychological continuity). Furthermore, when he wakes up, he is apparently bodily identical to how he was in the time before the time loop - he might have just been hit by a car in the loop, for instance, but shows no sign of physical injury. From the viewer’s perspective, outside the time-loop, he has mentally undergone a sudden and massive mental transformation. However, even with this huge shift, there is never any suggestion that he is anyone but Phil, because he is still the closest thing to Phil around. This counts against the mental continuity view as far as personal identity is concerned and towards an intuitively animalist and closest continuer view (and somewhat content similarity) version of personal identity.

---

[63] Interestingly, a swampman can only be identical in one way, but can be ‘close enough’ in numerous ways. For example, one of the ways a swampman may ‘get it wrong’ is by having memories of events tomorrow rather than today.

It may seem intuitively plausible that a swampman is more likely to have the correct memories up to this point than also of the future, but actually, it is orders of magnitude more likely that a swampman will have at least some correct future ‘memories’ than correct memories up to exactly the present point in time. Thus, Phil is far more likely to be a swampman on that particular basis.

Why do I say this? A swampman is unbelievably unlikely, but within this unlikeliness, it can have various kinds of mental states. Only one of these mental states is identical to the original it is ‘being the swamp-person of’. But in an enormous number more potential cases, it has mental states that the original person will have at some point in the future or did have at some point in the past.

An enormous number more again is mental states the original person has never had, nor will have. A subset of these is mental states the person would have if they were in a time loop under various future scenarios. A further subset, are states where that time-loop gives them correct beliefs about future events.

It may seem strange that a randomly generated being is more likely to have correct beliefs about future events than to have currently correct beliefs, but that is just the nature of logical space around these things.

People not in the time-loop, but in personal proximity to the time-looping person have no access to what may or may not have happened in the previous loop. All they experience is an abrupt change in the person's mental state and then ignore that abrupt change in terms of how they make their identity assignment.

How can we deal with cases where there is content similarity but not content identicality in terms of rights and privileges of individuals?

One outcome from the burglar case is that it shows that items and item identity are separate. Items exist in physical space, whereas item identities exist as mental concepts. The issue is that when trying to work out the facts about the identity of persons and items over time, where those items or persons are replaced with identical items or persons, we seem to get a divergent understanding of what is happening based on whether we are using content similarity or causal history to assess the situation.

However, the real issue is that we have a mental 'map' of the situation and adherence to this map is what we take to be the facts about identity rather than the physical identity as such.

The illustrative phrase is that "The map is not the territory", and in just such a way, the identity is not the item.

The identity is just something that we apply to the state of an object with a particular spatial continuity of its parts. Rather than identity just being defined in terms of causal history and content similarity, a functional definition would help map the intuitive identity.

Functional identity is just whatever successfully takes on the role of the item or person in question. This helps overcome a difficulty with content similarity that it can allow a person to swap in or out identical objects and persons within a network of expectations and entitlements but not create a new expectation or entitlement.

As I said earlier: "If the world ever returns to exactly the state that it was when that entitlement's creation, then that entitlement is in force again." However, this creates a problem where something comes into existence as a good example of that thing, or person, that *might* exist but is not a duplicate of any actual item or person. For instance, it seems like a swampman who is not a duplicate of someone cannot have a citizenship. If a swampperson slots into a particular role, then he or she can fulfil the personal identity of an existing person and take on their citizenship along with the rest of their identity, but if that person never existed to begin with, then the swampman's rights are severely curtailed, and that does not make much sense.

The basic problem is that the facts about the swampman cannot be contingent on existing facts about some other person. That would mean that the very same creature would have no rights at all if the original had not existed. There is a network identity answer to that problem, but also a narrow content answer.

The problem of differing swampman rights depending on a predecessor's existence gains greater clarity if we reset the swampman argument but under a different causal scenario:

There has been a case reported of a sheep giving birth to an apparently human-like figure, which local people took to be the result of a sexual liaison between a ewe and a human male.[64] This is somewhat more approaching the 'non-specific' swampman case, as people mistake a creature (a deformed lamb with a non-human cause) as human. So let us imagine a similar fictional case where I go swimming one day and drown, with my body sinking unrecovered into the depths of the sea. At the same time, a whale gives birth to a severely deformed whale calf that resembles me in every detail and has the same mental content I had at the point of my drowning. The calf, 'in its confusion,' swims away from its mother and 'strands' on the beach. Uniquely, the stranding is not fatal, and the calf makes its way to my house, puts on my clothes, and lives life as I did before.

In this case, if we respect content similarity above causal history as a criterion of value in what matters, we might well grant the calf the right to continue on my life, quite as if it had been the original 'me' that had returned from my swim. In this case, a narrow content identical being has inserted itself in my life-stream and is able to take advantage of the various emotional and legal linkages that apply to me and draw on my entitlements.

If we are to grant that case as one where the whale calf can take over my life, there is still the alternate case that such a calf is born, but I never existed. In that case, the calf may wade ashore as before, but are we then proposing that we should then identify it as merely a deformed whale calf and concerned citizens ought to attempt to push it back into the sea from whence it came to prevent its 'stranding'? That would seem to be too severe a call on causal history, but content similarity seems to have less to say about it, and this is because, in the first case, its rights are contingent on my original land-based existence. Since I do not exist for it to have content similarity with, it seems like it reverts back to being merely a deformed whale-calf, to be pushed back into the sea.

What needs to be the case, is that where such a creature emerges, that has no content similarity with an existing person, but in fact it has a plausible (but potentially entirely fictional) personal history, it is treated on the basis of its societal functional identity, not its factual and actual causal history. So, for instance, if it remembers going to a school that does not exist, but which would be a typical school and remembers a network of friends and acquaintances that do not exist, but are just the kind of people who could exist, so overall it fits well into the typical cultural setup of the era. In that case, it should be able to participate in that setup as one of its members.

The point is that it should not need to have *actual* content identity with an existing member of society in order to be able to partake in society, as long as its narrow content is within the range of the typical member's. This is a kind of content similarity that applies to cultural and national groups

---

[64] Pyatt, Jamie. "Horror as 'half human half-beast' is born to a sheep in South Africa", *news.com.au*, Web. http://www.news.com.au/technology/science/animals/horror-as-half-human-halfbeast-is-born-to-a-sheep-in-south-africa/news-story/69b3274687ff7fb34dc9d5daa4664335

rather than strictly individuals. In terms of the GHDS, if this was not the case, then whether the residents of Punxsutawney were even right to treat Phil as human or not throughout the GHDS would depend not only on the metaphysical setup, but also on his degree of identically to the original Phil. I have thus sought to point out why this is not the case. Whatever the origins of a time-traveller, swampman, or deformed whale calf, we ought (rationally) to treat them on the same moral and social basis as we would a being with the usual history.

Thesis Conclusion

We started by seeking the facts about personal identity. These facts turned out to be hard, if not impossible, to obtain. The reason for this is that people understand personal identity in a psychologically intuitive way. Making up thought experiments to probe that psychological understanding shows that it is more of a heuristic for our actual circumstances than a system that works in all possible cases.

A paradigm case is looking at a childhood photo and considering “whether or not it is me?” There may be no unique or ‘final’ facts that will make it me in all possible cases. There is no atomic threshold that makes it me or not me. It would be much easier to tell if I was an unchanging object, like a diamond floating through space. In this view, I am in agreement with Parfit.

In many ways, this result is disappointing as I had really wanted to find out the truth about personal identity and specifically use content similarity to solve the problem.

In terms of what matters in survival: This thesis bolsters the case against causality being what matters that Scott Campbell has already developed. What matters is content similarity, and that content similarity should be to as wide a scope as possible, so it includes our connections with as many other people as possible.

The critical achievement of this thesis is to prove that Parfit is incorrect when he says that what matters is only Relation R. Specifically, what matters in survival is Relation R and the ability to connect with other beings. I most convincingly set this out in chapter seven.

The concept of Network Identity is an important one. It provides a missing link and a further fact about what matters in survival. It shows that we already do not consider it to be survival where the connection element is absent. For example, we never thought there is survival in the case where someone is cut off from all access to the world, although many forgot this. Parfit was one of them, and his specific error was to see what mattered as psychological continuity and connectedness, whilst smuggling into that description the ability to connect with others, without acknowledging it.

In talking about Groundhog Day, we have the counterfactual circumstance to examine what really matters in survival, and we can see that what matters is Relation R plus connection with others.

There have been many twists and turns in this journey for me. Mostly it was finding again and again that Parfit was pretty much right about everything, but also finding that he was not clear on one big thing - causality, which Scott Campbell cleared up, and finally that he was specifically wrong in saying that all that mattered in survival was Relation R. I have made a huge amount of work to get there, but I am happy to have the success of this thesis judged primarily on the truth of this important philosophical theory.